HAL
open science

Planck intermediate results

Yashar Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, K. Benabed, et al.

To cite this version:

Yashar Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, et al.. Planck intermediate results: LV. Reliability and thermal properties of high-frequency sources in the Second Planck Catalogue of Compact Sources. Astronomy and Astrophysics - A&A, 2020, 644, pp.A99. 10.1051/0004-6361/201936794 . cea-03065753

HAL Id: cea-03065753

https://cea.hal.science/cea-03065753

Submitted on 14 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A&A 644, A99 (2020)
https://doi.org/10.1051/0004-6361/201936794
© ESO 2020

Astronomy & Astrophysics

Planck intermediate results

LV. Reliability and thermal properties of high-frequency sources in the Second *Planck* Catalogue of Compact Sources

Planck Collaboration: Y. Akrami[12,43,45], M. Ashdown[51,4], J. Aumont[74], C. Baccigalupi[62], M. Ballardini[17,32], A. J. Banday[74,8], R. B. Barreiro[47], N. Bartolo[20,48], S. Basak[67], K. Benabed[42,69], J.-P. Bernard[74,8], M. Bersanelli[23,36], P. Bielewicz[60,62], J. R. Bond[6], J. Borrill[10,72], F. R. Bouchet[42,69], C. Burigana[35,21,38], E. Calabrese[65], P. Carvalho[7,⋆], H. C. Chiang[19,5], B. P. Crill[49,9], F. Cuttaia[32], A. de Rosa[32], G. de Zotti[33], J. Delabrouille[2], J.-M. Delouis[52], E. Di Valentino[50], J. M. Diego[47], X. Dupac[26], S. Dusini[48], G. Efstathiou[51,44], F. Elsner[57], T. A. Enßlin[57], H. K. Eriksen[45], R. Fernandez-Cobos[47], F. Finelli[32,38], A. A. Fraisse[19], E. Franceschi[32], A. Frolov[68], S. Galeotta[34], K. Ganga[2], M. Gerbino[29], J. González-Nuevo[14], K. M. Górski[49,75], S. Gratton[51,44], A. Gruppuso[32,38], J. E. Gudmundsson[73,19], W. Handley[51,4], F. K. Hansen[45], D. Herranz[47], E. Hivon[42,69], M. Hobson[4], Z. Huang[66], W. C. Jones[19], E. Keihänen[18], R. Keskitalo[10], J. Kim[57], T. S. Kisner[55], N. Krachmalnicoff[62], M. Kunz[11,41,3], H. Kurki-Suonio[18,31], J.-M. Lamarre[70], A. Lasenby[4,51], M. Lattanzi[39,21], C. R. Lawrence[49], M. Le Jeune[2], F. Levrier[70], P. B. Lilje[45], V. Lindholm[18,31], M. López-Caniego[26,⋆], Y.-Z. Ma[61,64,59], J. F. Macías-Pérez[53], G. Maggio[34], N. Mandolesi[32,21], A. Marcos-Caballero[47], M. Maris[34], P. G. Martin[6], E. Martínez-González[47], S. Matarrese[20,48,28], N. Mauri[38], J. D. McEwen[58], M. Migliaccio[25,40], D. Molinari[21,32,39], A. Moneti[42,69], L. Montier[74,8], G. Morgante[32], P. Natoli[21,71,39], D. Paoletti[32,38], B. Partridge[30], F. Perrotta[62], V. Pettorino[1], F. Piacentini[22], G. Polenta[71], J.-L. Puget[41,42], J. P. Rachen[15], M. Reinecke[57], M. Remazeilles[50], A. Renzi[48], G. Rocha[49,9], G. Roudier[2,70,49], B. Ruiz-Granados[46,13], M. Savelainen[18,31,56], D. Scott[16], G. Sirri[38], L. D. Spencer[65], A.-S. Suur-Uski[18,31], J. A. Tauber[27,⋆], D. Tavagnacco[34,24], M. Tenti[37], L. Toffolatti[14,32], M. Tomasi[23,36], T. Trombetti[35,39], J. Valiviita[18,31], B. Van Tent[54], P. Vielva[47], F. Villa[32], I. K. Wehus[45], A. Zacchei[34], and A. Zonca[63]

(Affiliations can be found after the references)

Received 26 September 2019 / Accepted 18 July 2020

ABSTRACT

We describe an extension of the most recent version of the *Planck* Catalogue of Compact Sources (PCCS2), produced using a new multi-band Bayesian Extraction and Estimation Package (`BeeP`). `BeeP` assumes that the compact sources present in PCCS2 at 857 GHz have a dust-like spectral energy distribution (SED), which leads to emission at both lower and higher frequencies, and adjusts the parameters of the source and its SED to fit the emission observed in *Planck*'s three highest frequency channels at 353, 545, and 857 GHz, as well as the IRIS map at 3000 GHz. In order to reduce confusion regarding diffuse cirrus emission, `BeeP`'s data model includes a description of the background emission surrounding each source, and it adjusts the confidence in the source parameter extraction based on the statistical properties of the spatial distribution of the background emission. `BeeP` produces the following three new sets of parameters for each source: (a) fits to a modified blackbody (MBB) thermal emission model of the source; (b) SED-independent source flux densities at each frequency considered; and (c) fits to an MBB model of the background in which the source is embedded. `BeeP` also calculates, for each source, a reliability parameter, which takes into account confusion due to the surrounding cirrus. This parameter can be used to extract sub-samples of high-frequency sources with statistically well-understood properties. We define a high-reliability subset (`BeeP/base`), containing 26 083 sources (54.1% of the total PCCS2 catalogue), the majority of which have no information on reliability in the PCCS2. We describe the characteristics of this specific high-quality subset of PCCS2 and its validation against other data sets, specifically for: the sub-sample of PCCS2 located in low-cirrus areas; the *Planck* Catalogue of Galactic Cold Clumps; the *Herschel* GAMA15-field catalogue; and the temperature- and spectral-index-reconstructed dust maps obtained with *Planck*'s Generalized Needlet Internal Linear Combination method. The results of the `BeeP` extension of PCCS2, which are made publicly available via the *Planck* Legacy Archive, will enable the study of the thermal properties of well-defined samples of compact Galactic and extragalactic dusty sources.

Key words. catalogs – cosmology: observations – submillimeter: general

1. Introduction

The *Planck*[1] satellite (Planck Collaboration I 2016) was designed to image the temperature anisotropies of the cosmic microwave background (CMB) with a precision limited only by astrophysical foregrounds. To achieve its objectives, *Planck* observed the entire sky in nine broadband channels between 30 and 857 GHz. The *Planck* all-sky maps contain not only the CMB, but also a variety of diffuse sources of "foreground" emission – especially the Milky Way from radio to far-infrared wavelengths, as well as extragalactic backgrounds such as the cosmic infrared background (CIB) and Sunyaev–Zeldovich emission from clusters of galaxies. In addition to diffuse emission, the *Planck* maps contain emission from compact Galactic objects (cold dense clumps, supernova remnants, etc.) and a wide variety of unresolved external galaxies.

The *Planck* Catalogue of Compact Sources (PCCS; Planck Collaboration XXVIII 2014) contains compact sources

⋆ Corresponding authors:
P. Carvalho, e-mail: `f.pedro.carvalho@gmail.com`;
M. López-Caniego, e-mail: `mlopez@sciops.esa.int`;
J. A. Tauber, e-mail: `jtauber@cosmos.esa.int`

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

extracted from the *Planck* maps using the first 15 months of data. The source-detection algorithm was independent at each frequency and consequently the PCCS comprises nine independent lists. The second version of the catalogue (PCCS2; Planck Collaboration XXVI 2016) was produced using the full-mission data, obtained between 13 August 2009 and 3 August 2013.

At the frequencies observed by the High Frequency Instrument (HFI; 100–857 GHz), the diffuse sky background consists mainly of cirrus, i.e., dust emission from our own Galaxy, which covers a large part of the sky, is bright, and spatially fluctuates in a complex way (Low et al. 1984). The presence of this cirrus significantly complicates the detection and validation of compact sources, particularly because the statistical properties of this background are poorly understood, and since this cirrus contains localized structures that can be easily confused with genuinely compact sources. In addition, most of the compact sources expected in the frequency range 217–857 GHz, both Galactic and extra-galactic, have a dust-dominated spectrum similar to that of the cirrus.

The approach of PCCS2 to this problem was the cautious but simple one of defining a set of masks within which the cirrus emission was bright or complex, and labelling all compact sources found within these masks as "suspicious". The masks were derived at each frequency from: (a) brightness-thresholded total emission maps; and (b) maps of filamentary emission derived from a difference-of-Gaussians technique. All the compact sources detected in the union of these two masks were put into separate lists, referred to as PCCS2E (E for "Excluded"), and their reliability was not determined. The Exclusion masks include the Galactic plane and the low-Galactic-latitude regions, and cover from 15% of the sky at 100 GHz to 66% of the sky at 857 GHz. PCCS2E contains 2487 (43 290) sources at 100 GHz (857 GHz), to be compared to 1742 (4891) sources in the PCCS2 "proper"[2]. The vast majority of compact sources detected in the HFI maps therefore reside in the PCCS2E. While it is likely that many of the sources within PCCS2E are not genuine compact sources, but rather bumps or filaments in the cirrus background, inspection by eye of the maps clearly reveals that many of the sources are very probably genuine. Figure 1 shows a $10^\circ \times 10^\circ$ patch of sky on which the locations of both PCCS2 and PCCS2E sources are displayed. The lack of information on the reliability of the PCCS2E sources diminishes the overall utility of the PCCS2+2E. This new study addresses that problem.

We do this by making use of two kinds of information available in the *Planck* maps but not used by PCCS2. First, we use data from multiple frequencies simultaneously. The vast majority of high-frequency compact sources in PCCS2+2E, both Galactic and extragalactic, radiate thermal dust emission, which can be adequately modelled with a modified blackbody (MBB) spectral energy distribution (SED) characterized by a temperature and a spectral index (T, β). This smooth spectral behaviour can be used to improve the detectability and reliability of individual sources at high frequencies, while at the same time determining the parameters of the corresponding SEDs. This technique has been used to construct several previous *Planck* catalogues, including: the Catalogue of Galactic Cold Clumps (Planck Collaboration XXVIII 2016); the Catalogue of Sunyaev–Zeldovich Sources (Planck Collaboration XXVII 2016); the List of High-Redshift Source Candidates (Planck Collaboration Int. XXXIX 2016); the band-merged version of the Early Release Catalogue of Compact Sources (Chen et al. 2016); and the Multi-frequency Catalogue of Non-thermal Sources (Planck Collaboration Int. LIV 2018).

The second piece of information is that the brightness distribution of the diffuse cirrus emission varies relatively slowly and smoothly across the sky. This implies that its spatial-statistical properties are likely to be homogeneous within relatively large patches. In addition, since the cirrus itself has an SED of the MBB type, its spatial distribution is correlated across frequency channels. The statistical properties of the background can therefore be determined locally with good precision, and this information can be used to help separate sources from backgrounds.

We have carried out a re-analysis of all the sources contained in PCCS2+2E at 857 GHz[3], which assumes that a single compact source is responsible for the emission observed across a range of frequencies, both below and above 857 GHz. We further assume that each source can be distinguished from the diffuse background in which it is embedded, either by being an outlier (in the sense that its spatial distribution does not match the statistical properties of the background) or by exhibiting a significantly different SED. We combine multi-channel information re-extracted from *Planck* and IRAS maps to: (a) assess the reliability of detection of each source, taking into account potential confusion with the background; (b) re-determine the flux density of each source at frequencies from 353 to 857 GHz; (c) evaluate the spatial parameters (location and extension) of the compact source; and (d) estimate the parameters of an MBB fit to the emission across all the frequencies considered.

The results of this re-analysis are included in the *Planck* Legacy Archive[4] (PLA) as an extension of the PCCS2 and PCCS2E 857 GHz catalogues, appending the values of the new parameters to the original files. This extension of PCCS2 enables extraction of sub-samples that have well understood statistical properties, which in turn enables the study of the thermal properties of compact Galactic and extragalactic sources.

The outline for this paper is as follows. In Sect. 2, we present the data that we use as input to the analysis. In Sect. 3, we detail the model that we use to describe the sources and associated backgrounds, and we outline the Bayesian algorithm that we use to analyse each source and the main parameters that it outputs (details are given in Appendix A). In Sect. 4, we describe the simulations that we have built and used to tune and validate the algorithm and some of the main results. In Sect. 5, we describe how we produce and filter the new information added to the PCCS2+2E catalogue. In Sect. 6, we carry out a global characterization of the results of this analysis. In Sect. 7, we validate the results of this analysis against PCCS2 and other catalogues, and (for diffuse emission parameters) against dust maps derived from *Planck* data. In Sect. 8, we summarize our results, and provide recommendations for users of the new source information.

We have also included several appendices as follows. In Appendix A we detail the statistical machinery that we use. In Appendix B we describe how we have used our simulations to characterize and test the results. In Appendix C we comment on our Bayesian approach to contamination analysis, as opposed to a more classical frequentist approach. Finally, in Appendix D we include for reference the resulting SEDs that we obtain for a small number of well-known sources.

[2] In the rest of this paper we shall refer to PCCS2 as the list of sources *not* included in PCCS2E, and we shall call the union of both "PCCS2+2E".

[3] We have not attempted to re-detect and extract sources from the map, but instead use as starting point of our analysis the locations of all sources already existing in PCCS2+2E. However, all the photometric data used in our analyses are re-extracted from the *Planck* maps.

[4] https://www.cosmos.esa.int/web/planck/pla

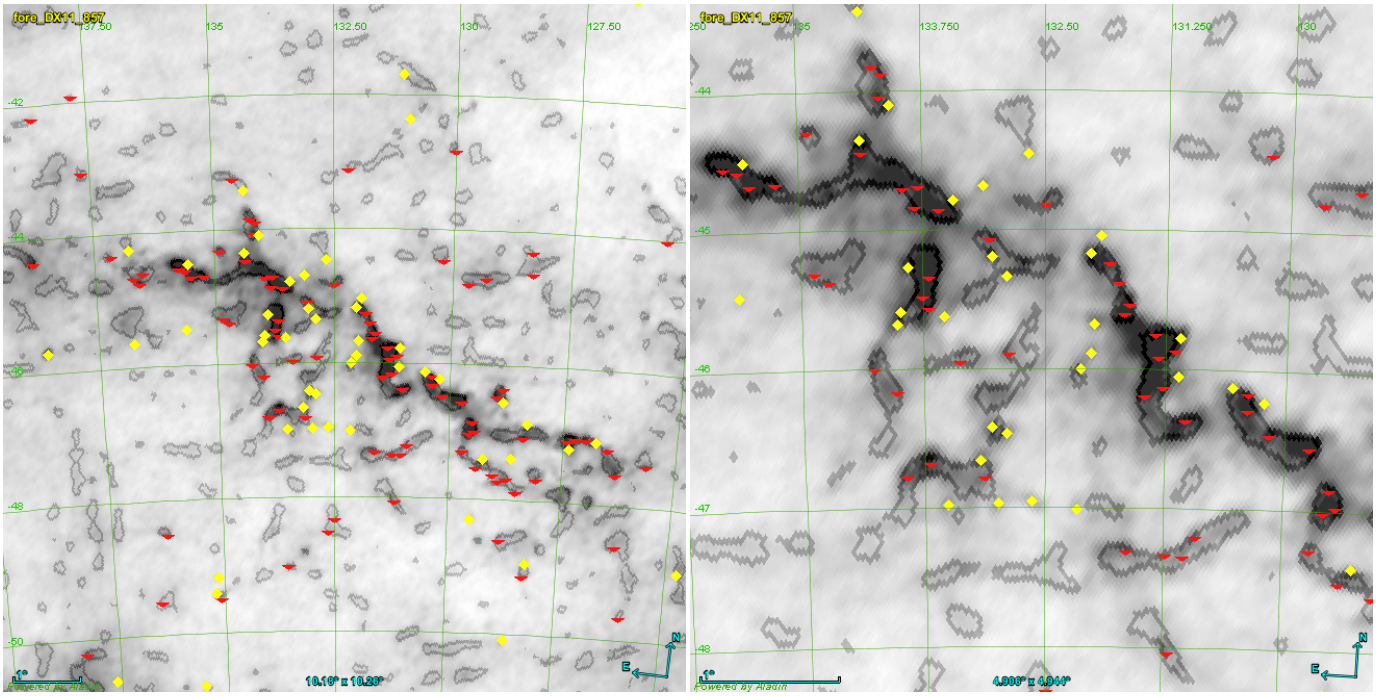

Fig. 1. *Left*: a 10° × 10° mid-Galactic-latitude ($b^{\rm II} \approx 45°$) region of the *Planck* 857 GHz map superimposed on the PCCS2+2E filament mask (grey contours). PCCS2 sources are yellow diamonds and PCCS2E sources are red triangles. The selected region contains complex backgrounds with localized features such as filaments and cirrus, causing the mask to break up into numerous islands. Many PCCS2+2E sources trace these structures, suggesting that some of the sources are parts of filamentary structures broken up by the source-finder and not genuine compact sources. *Right*: central 4.9° × 4.9° of the picture on the left, showing more clearly the spatial distribution of the PCCS2 and PCCS2E sources relative to the mask.

Parts of this paper describe details of our methods, and are necessarily long and technical. For readers whose main interests are the use of our results, we recommend to focus on Sects. 3.1 and 3.2, which describe our source and background models, and Sects. 5 and 6, which describe how we generate catalogue information, and how we then select a "base" catalogue of reliable sources. Section 7 compares our results to other catalogues, and can be skimmed unless such comparisons are important to the reader. Our main results are summarized in the final section, and Appendix D provides some specific examples of well-studied or interesting sources extracted from our catalogue.

2. Data

We use the 857 GHz source list of the Second *Planck* Catalogue of Compact Sources (Planck Collaboration XXVI 2016) to provide the initial source locations for our multifrequency Bayesian analysis. The angular resolution of *Planck* was highest at 857 GHz (corresponding to 4′.7), and this list contains the largest number of sources of any individual frequency in PCCS2. The 857 GHz source list contains flux densities for each source detected at 857 GHz, as well as estimates of flux densities at 545 and 353 GHz at the same locations. We note that the 857 GHz list does not contain any indication of the reliability of individual sources; the highest frequency at which such an indication is given is 353 GHz.

Our analysis then uses the *Planck* all-sky temperature maps at 353, 545, and 857 GHz from the *Planck* 2015 release (Planck Collaboration I 2016) to derive the characteristics of sources and their surrounding background. These maps are provided in the *Planck* Legacy Archive in HEALPix (Górski et al. 2005) format with $N_{\rm side} = 2048$. The description of these maps can be found in Planck Collaboration VII (2016). In addition, we use the 3000 GHz IRIS map, a reprocessed IRAS map described in Miville-Deschênes & Lagache (2005), with the same pixelization as the *Planck* maps[5].

Since the start of this work, a new generation of *Planck* maps has been released, which is referred to as the 2018 or Legacy release (Planck Collaboration I 2020). However, a new catalogue of compact sources has not been extracted from the Legacy maps. Therefore, we continue using the *Planck* 2015 maps that are the source of PCCS2.

3. Methodology

There is a long history of astronomers constructing catalogues, and many different approaches have been implemented, depending on the source and background properties. When the sources are unresolved and the background has no correlations, then the optimal approach is simply to use a point-spread-function filter (e.g., Stetson 1987) or thresholding methods appropriate for isolated sources, perhaps with varying noise levels, using software such as `SExtractor` (Bertin & Arnouts 1996). When the statistical properties of the background are known, one can instead use a matched-filter approach (e.g., Tegmark & de Oliveira-Costa 1998; Barreiro et al. 2003). If the background is more complex, if the sources themselves are partially resolved, or if the observed fields are crowded, the task of making a reliable catalogue becomes much more difficult. Several methods have been used to extract compact sources from confused Galactic regions, for example, using second derivatives and multi-Gaussian fitting as in `CuTEx` (Molinari et al. 2011), using higher-resolution

[5] For clarity, we do not use the DIRBE-inpainted maps which filled in the IRAS gaps.

data and multi-scale extraction as in `getdist` (Men'shchikov 2013) applied to *Herschel* data, a similar multi-scale approach with `Gaussclumps` applied to LABOCA data (Csengeri et al. 2014), or associating contiguous bright regions as a single source in `Clumpfind` (Williams et al. 1995) or `FellWalker` (e.g., Nettke et al. 2017) for SCUBA-2 data. A completely different strategy focuses on estimating the background properties simultaneously with the source properties, and that is the approach we follow here.

We carry out an independent Bayesian likelihood analysis (see e.g., Hobson et al. 2009) for each source contained in the 857 GHz catalogue of PCCS2+2E, and for the background surrounding it. The likelihood analysis takes as input four maps (353, 545, and 857 GHz from *Planck* 2015, and 3000 GHz from IRIS). We implement this analysis in software called the Bayesian Estimation and Extraction Package, and refer to it as `BeeP`. The analysis of each source assumes a model of the signal due to the source, and another due to the background.

3.1. Source model

We model the signal $\boldsymbol{s}_j$ due to the jth source as

$$\boldsymbol{s}_j(\boldsymbol{x};\boldsymbol{\Theta}_j) = A_j \boldsymbol{f}(\boldsymbol{\phi}_j)\boldsymbol{\tau}(\boldsymbol{x}-\boldsymbol{X}_j;\boldsymbol{a}_j), \tag{1}$$

where A_j is an overall amplitude for the source at some chosen reference frequency, which we take to be 857 GHz[6], $\boldsymbol{f}$ contains the emission coefficients at each frequency, which depend on the emission-law parameter vector $\boldsymbol{\phi}_j$ of the source (see below), and $\tau(\boldsymbol{x}-\boldsymbol{X}_j;\boldsymbol{a}_j)$ is the convolved spatial template at each frequency of a source centred at the position $\boldsymbol{X}_j \equiv \{X_j, Y_j\}$ and characterized by the shape parameter vector $\boldsymbol{a}_j$. Thus, the parameters to be determined for the jth source are its overall amplitude, position, shape, and emission law, which we denote collectively by $\boldsymbol{\Theta}_j = \{A_j, \boldsymbol{X}_j, \boldsymbol{a}_j, \boldsymbol{\phi}_j\}$.

If we make explicit the dependence of the source signal with the frequency channel (i), we have

$$\boldsymbol{s}_{ji}(\boldsymbol{x};\boldsymbol{\Theta}_j) = A_j f_i(\boldsymbol{\phi}_j)\left[\widehat{\boldsymbol{\tau}}(\boldsymbol{x}-\boldsymbol{X}_j;\boldsymbol{a}_j) * \boldsymbol{B}_i(x)\right], \tag{2}$$

where $\boldsymbol{B}_i(x)$ is the beam point-spread function of channel i. In this study we are mostly targeting completely unresolved objects, i.e., beam-shaped "point sources"; however, since PCCS2+2E also includes extended objects, we model the intrinsic shape of a source as a symmetrical two-dimensional Gaussian,

$$\widehat{\boldsymbol{\tau}}(\boldsymbol{x};\boldsymbol{a}\equiv r) \equiv \frac{1}{2\pi r^2}\exp\left(-\frac{x^2+y^2}{2r^2}\right), \tag{3}$$

where $\boldsymbol{a} \equiv r$ is the source radius.

The intrinsic spatial profile of the source $\widehat{\boldsymbol{\tau}}(\boldsymbol{x};\boldsymbol{a}_j)$ (before any instrumental distortion) is assumed to remain unchanged across frequencies[7]. To allow the intrinsic source size to vary with frequency would require more parameters and increased uncertainties to account for a situation that corresponds to a minority of sources. We have therefore chosen to impose a single, constant size parameter for a given source.

As mentioned in Sect. 1, the frequency spectra of most of the compact objects found in the *Planck*-HFI maps can be well-represented by an MBB spectrum (Planck Collaboration XXVI 2016); however, the SEDs of a minority of sources, for instance blazars, are not well-described by a modified blackbody. Therefore, we fit all sources with *both* MBB and "Free" models. In the latter, the emission coefficient f_{ν_i} at each channel is a free parameter. The MBB spectrum is written as

$$\ln f_\nu = \beta\,\ln\left(\frac{\nu}{\nu_0}\right) + \ln\left[\frac{B_\nu(T)}{B_{\nu_0}(T)}\right], \tag{4}$$

where the spectral parameters $\boldsymbol{\phi} = \{\beta, T\}$ are the dust emissivity spectral index and temperature, respectively, $B_\nu(T)$ is the Planck law of blackbody radiation, and ν_0 is once again the reference frequency. We normalize $\boldsymbol{f}$ so that $f_\nu = 1$ at $\nu = \nu_0$.

The Free model is written as

$$\boldsymbol{f} = [f_{\nu_1}, \cdots, f_{\nu_n}]^{\mathsf{T}}, \tag{5}$$

where the emission coefficients f_{ν_i} are free parameters. In effect, this model is a way to estimate source flux densities in each channel without imposing an SED, but still assuming that there is a single source at all frequencies. This extra flexibility comes at the cost of a larger model complexity, since it requires more free parameters. The flux-density estimates for the Free model are those that can most closely be compared to the ones already present in PCCS2+2E.

The location of the centre of the source is represented in Eq. (2) by $\boldsymbol{X}_j$. Our analysis initially assumes that the source is centred at the location defined in the 857 GHz list of PCCS2+2E. However, the source centre may be expected to vary slightly from channel to channel, and for this reason we allow our method to deviate from the initial values in an attempt to find the best overall location. Furthermore, during this investigation we realized that many of the source locations listed in PCCS2+2E are not well determined: in many cases we see that the centres of one or more sources are located around the edge of a well-defined blob of emission (e.g., Fig. 2). This problem affects about 10% of all sources in PCCS2+2E for the higher-frequency channels, and is inherent to the Mexican-hat wavelet 2 (MHW2) algorithm used to perform the detection. This wavelet, when used as a filter, is known to maximize the S/N of the objects, but it is also known to produce artefacts at a fixed distance from the centre of the source, Such artefacts related to the shape of the filter can be identified and removed particularly well in the cleaner regions of the sky. This additional cleaning step was performed for the lower-frequency channels of PCCS2, where the beamwidths are larger and these ringing effects are more prominent, but it was not performed for the higher-frequency channels because it was not considered necessary. Moreover, the MHW2 algorithm is well suited for the detection of point-like objects; however, when dealing with slightly extended structures such as those found at 857 GHz, the artefacts introduced by this filter are more evident, and a two-step cleaning procedure is definitely needed.

In our analysis we allow a new location to be determined from all the frequencies considered. As a result, in a number of cases several PCCS2+2E sources will be associated with the same physical source location[8]. However, there are also many genuinely independent sources that are relatively close to each

[6] The reference frequency does not need to be the centre of one of the data channels.

[7] The source shape is also convolved with the pixel window function at each frequency and this is taken into account in our analysis. In this particular case the pixel window function does not change across maps.

[8] There are 8269 (17.2%) PCCS2+2E locations that are associated with a different source. Of those 162 (3.3% of the PCCS2) are in the PCCS2 region.

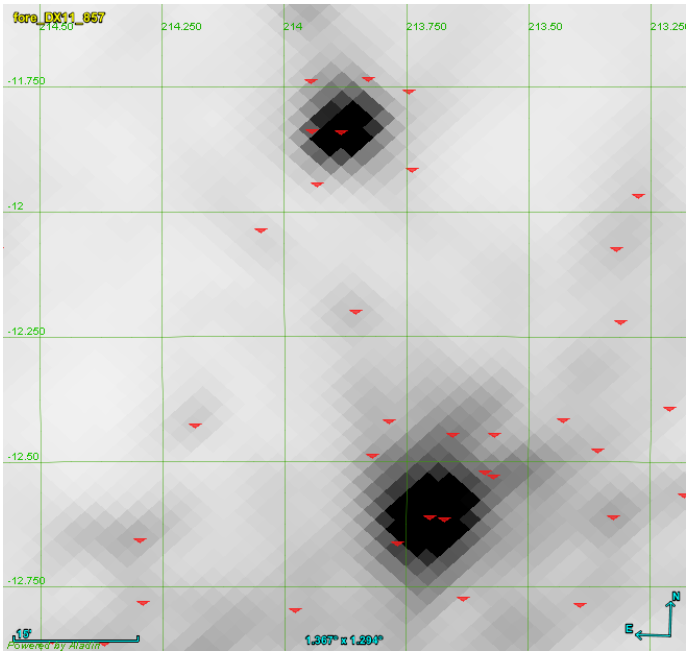

Fig. 2. Small patch (1°.3 × 1°.3) taken from the *Planck* 857 GHz map. The red triangles are PCCS2+2E catalogue positions. One can see that around the two bright "blobs" there are many associated PCCS2+2E sources, but their locations do not reflect the actual brightness peaks.

other, and there is a risk that the algorithm would "merge" them. We have therefore compromised by allowing our algorithm to move the location by at most 3 pixels (4′.5) away from its starting point. If this extreme is reached without an optimal solution being found, a flag indicating this is set in the final parameters.

3.2. Background model

We now need to account for the astronomical background $\boldsymbol{b}(x)$ and the instrumental noise $\boldsymbol{n}(x)$. A strong assumption of our framework is that the joint background in which the sources are immersed ($\boldsymbol{b}(x) + \boldsymbol{n}(x)$) is a two-dimensional, statistically isotropic Gaussian random field. Such a field is fully defined by its covariance matrix, which we use in our method as the mathematical representation of the background. The full-sky maps observed by *Planck* however, are neither statistically isotropic nor Gaussian. At high Galactic latitudes, the diffuse emission from Galactic dust is faint, and the (mostly extra-galactic) brighter compact sources stand out easily against it. However, the situation changes rapidly at low Galactic latitudes, as the diffuse emission competes in brightness with even the brightest compact sources. In this situation, confusion between "genuine" sources and the diffuse emission leads to difficulties in estimating the statistical properties of the background alone.

To improve our estimation of the properties of the background, we first reduce the size of the sky patch analysed around the source such that we can assume that statistical isotropy applies locally[9]. Second, we use the covariance matrix of the *cross*-power spectra across frequency channels. This improves the situation, since the instrumental noise $\boldsymbol{n}(x)$ is mostly uncorrelated across channels, and the astronomical background $\boldsymbol{b}(x)$ is better-determined by the larger data volume. The determination of an accurate cross-spectrum covariance matrix turns out to be a key element in our method. To improve the estimation of the off-diagonal components of this matrix, we filter out the noise component using the theory of random covariance matrices (Bouchaud & Potters 2004, Chap. 9). We have found that we also need to weight the off-diagonal elements (which represent the correlated part of the background) with respect to the diagonal elements (which represent the "noise") in order to accommodate the very large dynamic range of sources. The weighting factor that we use is tuned on simulations to reduce bias in the recovery of source parameters. More details on these analysis choices are described in Appendix A.

In practice, PCCS2+2E provides a list of *potentially* genuine sources that are embedded in the background whose properties we are estimating. For each of these sources, we create "background" maps (see Appendix A.2) by masking all surrounding PCCS2+2E sources[10] and inpainting the masked areas (see Sect. 5.3 of Casaponsa et al. 2013). We use a 7′ masking and inpainting radius to provide a good balance between effective source-brightness removal and preservation of the statistical properties of the background (see Figs. A.1 and A.2, and the discussion in Appendix A.1.4), especially at low Galactic latitudes where the density of sources is very high. Close to the Galactic plane, a large fraction of the background patch (up to 74% near the Galactic centre) is masked and inpainted, which might be expected to have a significant effect on the estimation of the detection significance[11]. More generally, we expect that inpainting may bias the estimation of the background properties, but it cannot be avoided because the effect of unremoved bright sources or of corresponding holes would certainly be much higher. The impact of inpainting cannot be modelled analytically, and the only way to assess it is through simulations. Simulations with different degrees of inpainting are discussed in Sect. 4, and show that the effect on source parameters is indeed small (as discussed further in Appendix B.2).

3.3. Combined model and its analysis

In this section we present the principles of our Bayesian analysis methodology. Appendix A gives technical details of the approach and its practicalities.

We first combine our models for sources and background into a model of the observed maps. A realistic model would have to include the entirety of sources and the full sky together. However, as described in Appendix A.1, under the assumption that the sources do not blend together, it is possible to simplify the problem and model each source independently:

$$\boldsymbol{d}_j(\boldsymbol{x}) = \boldsymbol{s}_j(\boldsymbol{x}; \boldsymbol{\Theta}_j) + \boldsymbol{b}_j(\boldsymbol{x}) + \boldsymbol{n}_j(\boldsymbol{x}), \tag{6}$$

where $\boldsymbol{d}_j$ is the data vector (pixel values), and $\boldsymbol{b}_j$ and $\boldsymbol{n}_j$ represent astrophysical and noise backgrounds in the neighbourhood of the source (j).

We can now build the likelihood of a single compact object as

$$\mathcal{L}(\boldsymbol{\Theta}) = \frac{\exp\left\{-\frac{1}{2}\left[\boldsymbol{d} - \widehat{\boldsymbol{b}} - \boldsymbol{s}(\boldsymbol{\Theta})\right]^{\mathsf{T}} \mathsf{N}^{-1} \left[\boldsymbol{d} - \widehat{\boldsymbol{b}} - \boldsymbol{s}(\boldsymbol{\Theta})\right]\right\}}{(2\pi)^{N_{\rm pix}/2}\, |\mathsf{N}|^{1/2}}, \tag{7}$$

where $\widehat{\boldsymbol{b}}$ is the generalized background ($\boldsymbol{b} + \boldsymbol{n}$), N is the generalized background covariance matrix, and all individual source

[9] The "field" size we select is 3°.69 × 3°.69. The motivation for this choice can be found in Appendix B.1.

[10] For this purpose we merge all three source lists between 353 and 857 GHz.

[11] The fraction of inpainted pixels in the patch is reported in one of the columns of the catalogue and can be used to filter the selection.

parameters have been concatenated into $\boldsymbol{\Theta}$ for convenience. For clarity we have dropped the source index here.

The above expression allows us to consider the likelihood of a "no-source" model $\mathcal{L}_0$, when A, the source amplitude is 0. $\mathcal{L}_0$ is a constant, since it does not contain any parameters. The expression that we seek to maximize is the log of the $\mathcal{L}(\boldsymbol{\Theta})/\mathcal{L}_0$ ratio, which represents the likelihood that there is a source in addition to the background.

If $\widehat{\boldsymbol{\Theta}}$ is the parameter set that maximizes the likelihood ratio (Eq. (7)), then we define the quantity $\mathcal{R}$, corresponding to NPSNR in the catalogue[12] (the Neyman–Pearson signal-to-noise ratio), by

$$\ln\left(\frac{\mathcal{L}(\widehat{\boldsymbol{\Theta}})}{\mathcal{L}_0}\right) = \tfrac{1}{2}\mathcal{R}^2. \tag{8}$$

$\mathcal{R}$ is the detection significance level that expresses the number of sigmas of the detection, and is given in the NPSNR column of the BeeP catalogues. In the case that all of our assumptions hold, and all source parameters are known except amplitude, A, then $\mathcal{R}$ would in fact be the inverse of the fractional error on the amplitude, $A/\Delta A$. However, in practice, as we shall see, typical values of $\mathcal{R}$ are considerably higher than $A/\Delta A$. This is the result of either broken assumptions or uncertainties on the other estimated parameters that propagate into the source amplitude. In particular, the presence of cirrus produces strong positive-tail events in the likelihood, and this might be interpreted (erroneously) as generated by the source of interest (see Fig. 3 for examples).

To account for this effect, we build an estimate of the non-Gaussianity of the background that is independent of the likelihood, which we refer to as RELTH. Essentially we look in the background patch for outliers to a white-noise, unitary ($\sigma = 1$) Gaussian random field in pixel space ($\boldsymbol{X}$), which is what we would expect if all our assumptions hold, in other words, under the null hypothesis of our model. We assume that the positive outlier pixels created by the source itself are no more than a small fraction of the total number of pixels in a small patch around the source. Using the definition of quantiles, one would expect that

$$\int_{-\infty}^{\mathsf{RELTH}} \frac{\exp\left[-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right]}{\sqrt{2\pi}\sigma}\mathrm{d}x = 1-\alpha, \tag{9}$$

where RELTH is the $1-\alpha$ distribution quantile, and σ is the width of the Gaussian. Using simulations, we have verified that the fraction of outlier pixels created by the source is less than 5% of the total, so we use $\alpha = 5\%$.

RELTH can be read directly from the histogram of the actual field, and then Eq. (9) solved for σ. If the background pixels ($[1-\alpha]\%$ of the patch pixels) comply with the assumptions of the background model, then they will follow a unitary Gaussian distribution and the solution of Eq. (9) is $\sigma = 1$. However, as a result of the intrinsic non-Gaussianity of the background, the tails of the background histogram are expected to be larger than those of the unitary Gaussian distribution. This distribution of background pixel brightness with extended tails can then be approximated by a Gaussian, but with $\sigma > 1$ to account for the larger tails. Solving Eq. (9),

$$\sigma = k\ \mathsf{RELTH}, \tag{10}$$

where k is a pure numerical constant given by

$$k = \frac{1}{\sqrt{2}\,\mathrm{erfc}^{-1}(2\alpha)}, \tag{11}$$

[12] Identifiers in sans-serif capital letters correspond to column labels in BeeP output catalogues.

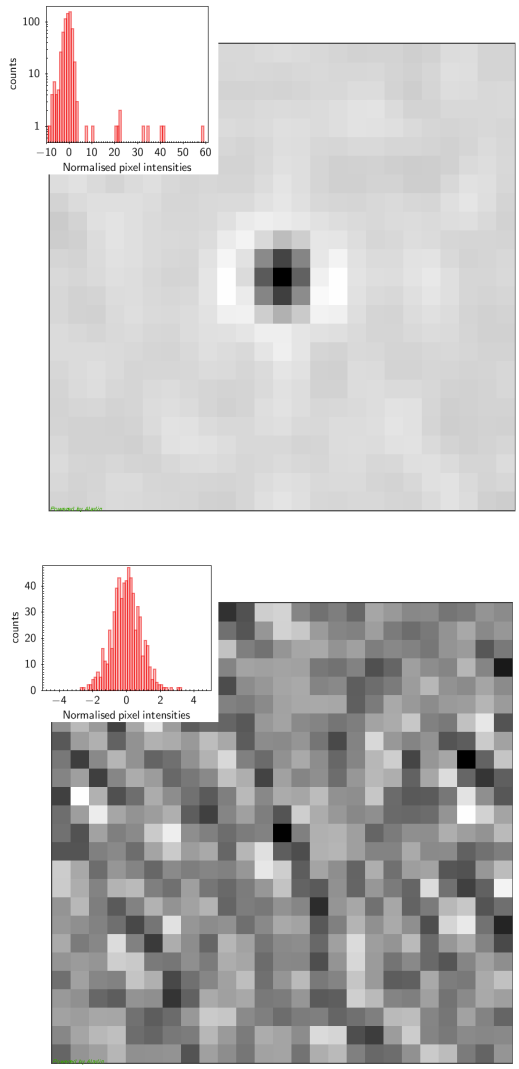

Fig. 3. Examples of potential analysis fields. *Upper panel*: high significance source (PCCS2 857 G172.20+32.04). The histogram (shown in the inset; Y-scale is log) is a mixture of a Gaussian component from the background pixels, plus a strong upper tail generated by the source in the centre. *Lower panel*: field with no detected sources in it (PCCS2 857 G172.20+32.04; Y-scale is linear). This time only the Gaussian component is present. The tails of the distribution are compatible with "just background". Each field is 25×25 pixels (1 pixel 1.′717). The pixel intensities (Eq. (A.13)) are unitless.

and erfc^{-1} is the inverse complementary error function.

We can now correct our "naive" significance NPSNR and define a new source significance variable as

$$\mathsf{SRCSIG} = \frac{1}{k}\,\frac{\mathsf{NPSNR}}{\mathsf{RELTH}}, \tag{12}$$

where k is a constant given by Eq. (11), which is the same for all sources. SRCSIG expresses the likelihood that there is a source in the patch being analysed. If the histogram of the background patch is Gaussian, then $\sqrt{2}\,\mathrm{erfc}^{-1}(2\alpha) = \mathsf{RELTH}$ by definition and SRCSIG = NPSNR. If our initial assumptions hold, as predicted, then NPSNR is the detection significance. However when there is non-Gaussianity in the background, either from diffuse components or localized features, then RELTH increases and a penalty is applied to the Gaussian criterion. The penalty is reduced towards high Galactic latitudes away from cirrus, where

the isotropy and Gaussian assumptions hold well. In the neighbourhood of the Galactic plane, or inside cirrus structures, the criterion becomes more stringent in order to avoid false positives induced by the non-Gaussianity of the background[13].

Finally, we note that RELTH depends on the detailed statistics of the field brightness. Therefore its ability to provide an estimate of the relative level of non-Gaussianity in the background is not uniform across the sky. However, tests using simulations show that it is effective both at low and high Galactic latitudes, and it can safely be used to correct NPSNR. On the other hand, it should probably not be used to directly compare levels of non-Gaussianity in regions that differ significantly in complexity.

4. Simulations

We have tested our method extensively using simulations. These tests have allowed us to tune parameters intrinsic to the method, and to assess the quality of the extracted source descriptors. There are four types of simulations, as follows.

1. *Synthetic* simulations (Appendix B.1) comprise data that mimic a basic assumption of the method as closely as possible, namely that the background is a homogeneous Gaussian random process. To make these simulations, we combine CMB map realizations based on the *Planck* 2015 best-fit cosmological model, with noise consistent with that of the *Planck* detectors as described in Planck Collaboration XII (2016). To these we add Gaussian sources whose thermal emission characteristics are taken from a preliminary BeeP extraction. We use these simulations to test the algorithm, and fix some of its basic parameters, such as the optimal size of the patch analysed around each source, and to check the impact of some systematics such as projection distortions.
2. *Injection* simulations (Appendix B.2) attempt to reproduce the properties of the diffuse backgrounds that are seen by *Planck*. The basic principle is to use the 2015 *Planck* maps and add to them a known set of sources. We have produced three distinct types of these simulations: (a) we remove from the observed maps the sources present in PCCS2+2E, inpaint the holes, and inject at the same locations point-like sources whose thermal emission parameters are those of the original source (as extracted by BeeP in a preliminary run); (b) as in (a), but the fake sources are injected in the vicinity of the original ones rather than at the PCCS2+2E location; and (c) the locations of the fake sources are randomly drawn from a uniform distribution over the high-latitude sky, and their thermal properties are drawn from the distribution present in PCCS2. In this case the original PCCS2+2E sources are not removed from the maps. In addition, we have also produced realizations of the above three types that include known source extensions. As detailed further in Appendix B.2, these simulations allow us to:
   - assess the effect of inpainting on the results;
   - determine an optimal level for the covariance matrix cross-correlation factor;
   - assess biases in the recovered source parameters, e.g., temperature and spectral index;
   - assess the accuracy of the estimated source locations, and on this basis establish a correction to the estimated location uncertainties; and
   - assess biases and establish corrections to both the estimated flux densities and their uncertainties (see Sect. 6.2.4).
3. *FFP8* simulations (Appendix B.4) are the most realistic realizations of the all-sky maps as observed by *Planck* and processed through the PR2 pipelines[14], and are fully independent of the observed maps. In particular they reproduce the variation across the sky and in frequency of the *Planck* beams, which is something that we do not include in our injection simulations. However, an important drawback is that a corresponding simulation of the IRIS sky is not available and therefore we cannot extract thermal-emission parameters in order to compare them directly to BeeP's results on *Planck* maps. Nonetheless, we are able to use these simulations to assess the impact of the beam variation on the recovery of flux densities and on the positional error, and on this basis we establish a correction to the flux-density estimates.
4. *No-source* simulations (see Sect. 5.1) use a list of locations that are *not* present in PCCS2+2E, and on which we run BeeP. Under the assumption that such locations contain only background emission[15], these simulations allow us to estimate the number of spurious sources generated by BeeP, i.e., the background-related contamination fraction of the resulting catalogue. The empty locations are selected in the neighbourhood of the catalogue positions in order to preserve the distribution of sources on the sky. We have placed the sources at a random location within an annulus of radii 12′ and 14′, enforcing that each injection location is at least 12′ from any other. We then mask and inpaint the original source.

All of the above tests and their results are described in detail in the Sects. 5.1 and 5.2, as well as Appendix B.

5. Catalogue production

The basic principles of the production methodology for the catalogue are described in Sect. 3 and implementation details in Appendix A. The BeeP software takes as input a catalogue of sources and associated maps, and processes all sources. The output is an extension of the input catalogue, in effect adding to each source a number of new parameter fields.

As described in Sect. 2, the input catalogue is the union of the 857 GHz PCCS2 and PCCS2E (PCCS2+2E) source lists, which contains 48 181 entries. The input data are the 2015 *Planck* full-mission frequency maps between 353 and 857 GHz, and the IRIS map. The IRIS map does not cover the full sky, and therefore a small subset of sources (650) has been processed with *Planck* channels only. This restriction seriously impairs the constraining capabilities of the likelihood, and hence a downgraded quality status has been assigned to these sources. As a consequence, the output catalogue contains 47 531 complete entries. Of those, 42 869 (about 90%) are in the PCCS2E, and only 4662 (10%) in the PCCS2.

5.1. Reliability assessment

Once we have processed the entire input catalogue through BeeP, we can apply filters to select subsets of sources. The first and most critical filter is reliability. For this purpose, we

[13] See Appendix A.1.

[14] The PCCS2+2E source catalogues were produced from the PR2 maps. The newer PR3 maps released by *Planck* in 2018 are based on significantly different pipelines, and have not been used to generate source catalogues; for this reason we cannot use the newest FFP10 simulations associated with PR3.

[15] This assumes that for the level of sensitivity we are aiming at, the PCCS2+2E catalogues are almost complete (Planck Collaboration XXVI 2016).

interpret our detection significance statistic SRCSIG in terms of reliability.

In a classical frequentist framework, we would draw the test receiver operational characteristic curve (ROC, Trees 2001, Chap. 2). The ROC curve shows the balance between "completeness", or true positive rate, and the false positive or "spurious" rate, when varying the threshold of the detection significance statistic. However, since we are not adding any new entries to the PCCS2+2E catalogue, we will always be limited by the initial catalogue's completeness. Our focus will therefore be on the spurious error rate or "contamination". The spurious error rate is the probability of classifying a source as real when only background is present for a given SRCSIG[16],

$$\text{Contamination} \equiv \Pr\left(\text{real} \,|\, \text{only background}; \mathsf{SRCSIG}\right). \tag{13}$$

Owing to the complexity of the data, the most practical way of estimating contamination is through simulations. For this purpose, we use the no-source simulations described in Sect. 4. We run the BeeP algorithm on the no-source catalogues, and compute the SRCSIG statistic. We then compute the percentage of locations where there is not a source for which the SRCSIG statistic is larger than a certain threshold. This gives an estimate of the contamination (Eq. (13))[17], under the assumption that there are no sources.

Figure 4 shows how this estimate of the contamination varies with the SRCSIG threshold, for two different thresholds of NPSNR. Solid blue lines are full-sky results, and dashed lines correspond to a catalogue restricted to PCCS2 sources. The solid (full-sky) green line is obtained similarly, but the original source is not removed. This test is carried out to show that the presence of the original source in the background significantly and systematically modifies the non-Gaussianity of the background in the area being analysed, reducing in a systematic way the SRCSIG distribution. As can be seen in Fig. 4, this effect would artificially (and incorrectly) reduce the contamination for a given SRCSIG threshold.

Figure 4 shows that if the catalogue is restricted to the more reliable sources, there is very little difference in the contamination levels of the PCCS2+2E full catalogue (solid line) and the PCCS2 subset (dashed line); this indicates that BeeP accounts adequately for the non-Gaussianity of the background. We select SRCSIG > 3.7 as an interesting threshold, which leads to a contamination level between 5% and 10% (Fig. 4).

Our simulation-based estimate of contamination relies on the prior assumption that there are no sources at the locations analysed, which is probably not correct for PCCS2+2E where crowding becomes significant. This makes the estimate of Eq. (13) a conservative one. The curves in Fig. 4 should then be read as the maximum contamination for a given SRCSIG threshold. To make it more realistic, the estimate should be reduced taking into account the catalogue completeness, as described in Appendix C. However, for high values of NPSNR, the correction is very small[18]; in this case one can safely use Fig. 4 as a reasonable estimate of the catalogue contamination. Comparison of the solid and dashed lines in Fig. 4 also shows the effect of crowding on contamination, which is at most 10% for low SRCSIG.

With the above considerations, a catalogue can be selected to have a given reliability level by adopting thresholds in SRCSIG and NPSNR. For example, if we define the condition

$$\mathsf{SRCSIG} > 3.7 \wedge \mathsf{NPSNR} > 5.0, \tag{14}$$

where the symbol "∧" means "logical and", the resulting catalogue has a maximum contamination between 5 and 10%[19]. The reliability condition in Eq. (14) is one of the important components for building the "BeeP/base" catalogue (see Sect. 5.5).

5.2. Rejection of outliers

As a result of the large range of source flux densities and the background conditions, it is reasonable to expect that under extreme conditions the simplified data model, and the likelihood, become a sub-optimal description of the statistical properties of the data, and that significant outliers will arise. As one of our goals is to have a well-defined set of statistical descriptors for the catalogue estimates, these extreme outliers need to be identified and removed to avoid biasing or distorting the characterization.

The extensive set of simulations described in Sect. 4 was used to identify such cases (see Appendix B.2 for more details). We find that any sources whose estimates do *not* meet the following "outlier-rejection criterion" must be considered unreliable:

$$\begin{aligned}\mathsf{EXT} > 1.46 \wedge \mathsf{TEMP} < 60 \wedge \mathsf{BETA} < 5 \\ \wedge\, (\mathsf{TH2SB} - \mathsf{TL2SB}) > 0.8 \\ \wedge\, (\mathsf{BETAH2SB} - \mathsf{BETAL2SB}) > 0.25,\end{aligned} \tag{15}$$

where EXT, TEMP, and BETA are the estimated source extension, temperature, and spectral index, respectively. The differences (TH2SB − TL2SB) and (BETAH2SB − BETAL2SB) are the estimated uncertainties of the temperature and spectral index[20]. The value of EXT that we use to create the filter is the "uncorrected" source size parameter (see Sect. 6.2.2 and Appendices A.2.3, A.2.4).

The criterion of Eq. (15) selects a very small fraction of the catalogue sources (2462, or about 5%). Of those, 1463 would also have been rejected by the reliability criterion (Eq. (14)). Thus only 999 or 2% of the sources that pass the reliability criterion are rejected by the outlier-rejection criterion (Eq. (15)).

5.3. Convergence filter

Our logical framework assumes a binary classification scheme, such that each region of interest is either diffuse background or a compact source. However, a binary classification model, regardless of the significant advantage of its simplicity, is not complete enough to explain the full complexity of the data set. In fact, as described in Sect. 5.1 (see also Appendix C), we

[16] In Appendix C we present the procedure using the "dialect" of the orthodox hypothesis testing framework.

[17] The uncertainty in the contamination estimate is $\sqrt{p(1-p)/(n+3)}$, where p is the contamination and n is the number of "false sources" ($n \approx 20\,000$). Even for large n, as in our case, some care must be used when selecting very low contaminations. An estimated contamination of 0.005 already carries an uncertainty of about 10%.

[18] Indeed, by comparing the curves with the two NPSNR thresholds shown in the left and right panels of Fig. 4, it can be deduced that the correction must already be very small at NPSNR > 3, a very low value for NPSNR.

[19] If we did not impose NPSNR > 5, then SRCSIG > 3.7 alone would set contamination to approximately 10%.

[20] We reject sources where the recovered parameter uncertainties are extremely low, indicating that the likelihood sampler has not been able to explore the parameter space adequately. There may be some exceptional cases where the uncertainties are very low because the model fits the data extremely well, and these will also be rejected. One such example can be seen in Fig. D.2. It is possible, by examining the results of BeeP, especially the χ^2 of the free model fit, to decide that the case should not be rejected.

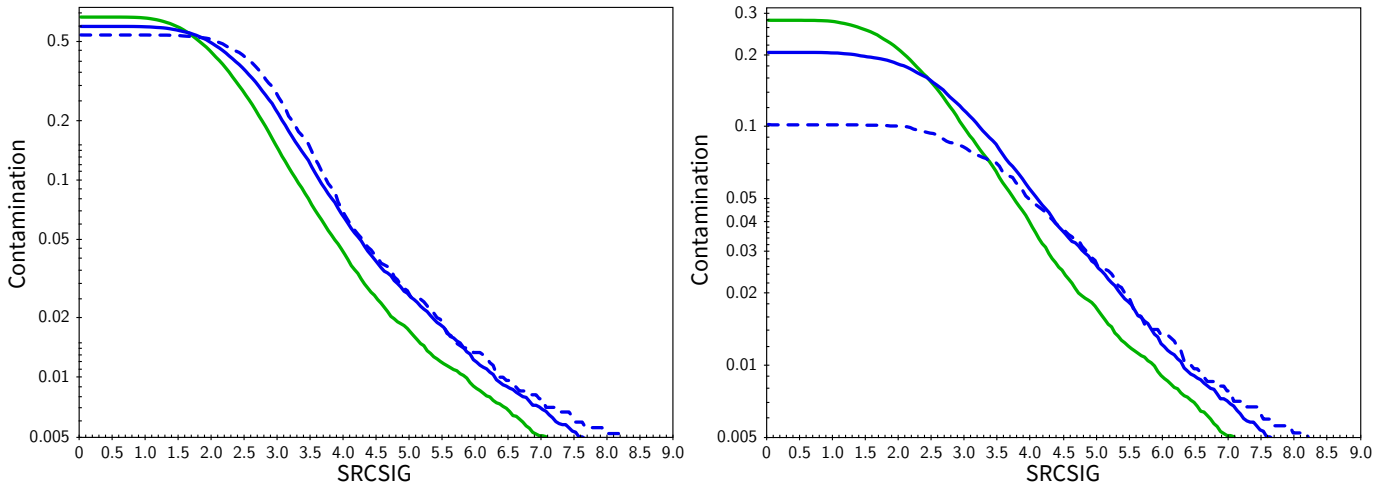

Fig. 4. Contamination (Eq. (13) and Appendix C) when SRCSIG $\geq x$, in the cases of NPSNR > 3 (*left*) and NPSNR > 5 (*right*). The blue curves, solid for full sky (PCCS2+2E) and dashed for high Galactic latitudes (PCCS2), display the contamination for simulations when the "no sources" are located in the neighbourhood of actual catalogue positions. The original sources are masked and inpainted. The solid green line shows the estimated contamination of a simulation exactly like that of the solid blue curve (full sky), but this time the original sources are not removed and inpainted. In this figure and several others in this paper, we label the axis with the name of the corresponding field in the output of BeeP analysis (in capital roman letters), e.g., here "SRCSIG".

compute the probability of a set of pixels *not* being part of the diffuse background (rejection of the null hypothesis), and the SRCSIG statistic acts as the discriminating variable. This mathematical machinery requires us to find a likelihood maximum in the proximity of the source position. However, in some cases, e.g., at low Galactic latitudes or along very extended sources, that condition may not be met. For example, in Fig. 5 there are some PCCS2E positions (blue triangles) that are well separated from the actual centre of the compact object, which coincides with the likelihood maximum. Since we have limited the likelihood "travel" distance to three pixels from the original PCCS2E+2E location (see Sect. 3.1 and Appendix A.2), in some of these cases BeeP fails to find a maximum. The code then assumes that the original PCCS2+2E position is correct, and samples the likelihood field around it. For extended sources where BeeP could not find a likelihood maximum, such as those shown in Fig. 5, SRCSIG can still attain a high value because the location does not have background-like properties. For this reason we have introduced a new catalogue field MAXFOUND, that flags when a likelihood maximum was found. Considering that being above a given SRCSIG threshold means, it is likely that this is not part of the background. MAXFOUND then allows one to discriminate between a compact object (value 1, Fig. 5, green squares) or something else (value 0, Fig. 5, red squares).

In Fig. 6 we show the total fraction of PCCS2+2E sources with NPSNR > 5 and above a given SRCSIG. The dashed curves in Fig. 6 show the impact of adding the condition of MAXFOUND = 1. The intersection of the curves with the SRCSIG = 0 axis shows the fraction of sources with NPSNR > 5.

5.4. Quality filter

We summarize the quality of the source parameter estimates in a new field, EST_QUALITY, which assigns five points to each source and subtracts penalties from this maximum value if certain quality criteria are not met. EST_QUALITY = 5 means that the estimates of source parameters are highly reliable. Penalties subtracted if specific quality criteria are not met

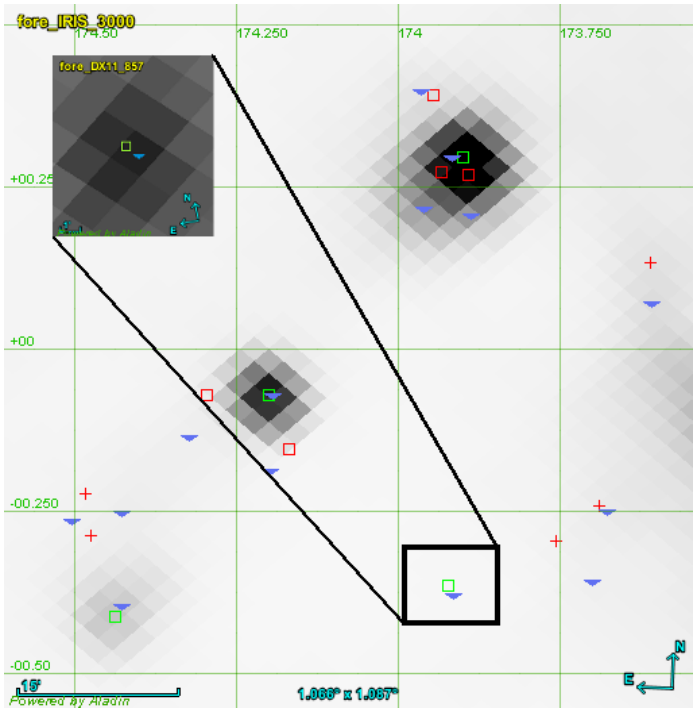

Fig. 5. Patch of 1°.0 × 1°.0 area centred on l = 173°.91, b = +00°.25, from the IRIS 3000 GHz map.

are listed in Table 1. When MAXFOUND ≠ 1 (no likelihood maximum), it is not possible to guarantee an optimal extraction of source parameter estimates. However, sources that fail only the MAXFOUND condition may still be used in many cases where a rigorous statistical characterization is not required. For this reason the associated penalty was set to half of the other criteria. Source estimates not meeting the "outliers criterion", or that were examined in only the *Planck* channels (because they are located in the IRAS gaps), should be used with great caution.

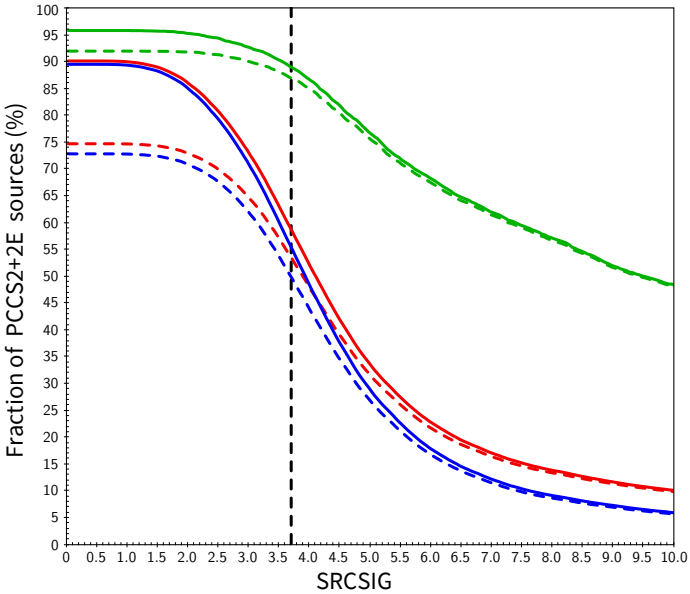

Fig. 6. Fraction of PCCS2+2E sources with NPSNR > 5 and above a given SRCSIG threshold. Green curves show PCCS2 sources, blue curves show PCCS2E sources, and red curves show the full PCCS2+2E. Dashed lines are the result of imposing MAXFOUND = 1. The dashed black vertical line (SRCSIG = 3.7) is the reliability criterion threshold that we have selected for the BeeP/base catalogue.

Table 1. Penalties applied to sources whose parameter estimates do *not* meet the quality criteria (note that the maximum quality level is 5).

Sources *not* meeting:	Penalty
EXT > 1.46 ∧ TEMP < 60 ∧ BETA < 5 ∧ (TH2SB – TL2SB) > 0.8 ∧ (BETAH2SB – BETAL2SB) > 0.25 (outliers, Eq. (15))	−2
MAXFOUND = 1 (no likelihood maximum)	−1
All four channels used (IRIS data missing)	−2

5.5. BeeP/base catalogue

Let us now examine the sub-catalogue defined by the conditions given in Eq. (14). If we require EST_QUALITY ≥ 4, this sub-catalogue contains 24 511 of the 43 290 objects in the PCCS2E (56.6%). If we require EST_QUALITY = 5, however, we still find 21 997 sources (50.8% of the PCCS2E objects). We therefore add this condition and define a "reliable and accurate" sub-catalogue based on the three following conditions:

$$\mathrm{NPSNR} > 5 \;\wedge\; \mathrm{SRCSIG} > 3.7 \;\wedge\; \mathrm{EST_QUALITY} = 5. \qquad (16)$$

This sub-catalogue, which we shall refer to as BeeP/base, contains 26 083 (54.1% of the full PCCS2+2E) objects. Unless otherwise stated, all figures in the rest of this paper are based on it. If we require a more stringent contamination level, say below 1%, (SRCSIG > 7.0 and EST_QUALITY = 5), there remain 5077 (11.7%) compact objects in the PCCS2E.

Although in the PCCS2+2E there is no indication of the source-detection significance, for comparison we computed one by dividing the MHW2 estimates of the source flux density and its uncertainty, DETFLUX/DETFLUX_ERR. The median value of the PCCS2+2E-estimated S/N (8.96) is considerably lower than the equivalent value of NPSNR in the BeeP catalogue (12.82). However, one must remember that BeeP is a multi-channel method, and jointly analysing more than one frequency strengthens the background-rejection criterion.

5.6. Beyond BeeP/base

In Sect. 5.5 we have described how we have extracted a subset of the sources in PCCS2+2E (BeeP/base) that we consider to be "reliable and accurate". Based on our analysis, this means that:
- the uncertainties on the extracted model parameters are realistic;
- the number of false detections is low.

We caution the user of BeeP/base that the parameter uncertainties for many sources in this catalogue are relatively large. For example, Fig. 11 (supported by simulations in Appendix B, see e.g., Fig. B.7) shows that sources with the lowest NPSNRs have flux-density extraction uncertainties larger than about 40%. At first glance this does not seem consistent with a naive interpretation of NPSNR as an "SNR-like" quantity, but we remind the reader that NPSNR reflects the uncertainties of *all* model parameters, not only flux-density determination. Figure 8 shows in particular that the flux-density determination is correlated with other parameters (in particular the size, temperature, and spectral index), and this certainly contributes significantly to increasing the uncertainties.

We have selected BeeP/base as a good approach for studying the broad characteristics of the results of our analysis. However, we expect that each user of these results will select a specific subset of sources based on their own needs. For example, if low flux-extraction uncertainties are required, then the threshold on NPSNR should be correspondingly increased, and we suggest using Fig. 11 as a guideline. Similarly, Fig. 4 can be used to set a threshold related to contamination by false detections. Each user of our results should determine the specific criteria that need to be applied to meet their objectives.

6. Base catalogue characteristics

We now describe and characterize the BeeP/base catalogue. As mentioned previously, all the results of this analysis (i.e., for all PCCS2+2E sources, not only those in BeeP/base) are available online via the Planck Legacy Archive. The Explanatory Supplement (Planck Collaboration ES 2018), which accompanies the results, includes an annotated list of all the parameters provided for each source. In this paper, we provide a summary of the key parameters in Table 2. Some of these are described in more detail in this section.

6.1. Reliability and quality parameters

The set of reliability and quality parameters includes:
- NPSNR, which measures the S/N of the combined detection (Eq. (8));
- SRCSIG, which measures the likelihood that the source is a real compact object distinct from the background (Eq. (12));
- EST_QUALITY, which measures the trustworthiness of the source descriptor estimates extracted by BeeP (see Sect. 5.4).

It is important not to confuse the roles of SRCSIG and EST_QUALITY. SRCSIG indicates the likelihood of a source being real, whereas EST_QUALITY provides an assessment of the quality of the estimated source parameters, given that the source is real. For instance, a bright nearby object may have a very large SRCSIG because we are sure it is a real object. Nonetheless it might still fail the EST_QUALITY criteria if,

Table 2. Summary of the key parameters generated by BeeP for each source in PCCS2+2E and available online via the *Planck* Legacy Archive.

Component	Extracted parameters
Source	New location
(thermal model)	Extension
	Thermal SED properties
	[Temperature, Spectral index, Ref. Flux Density]
	Flux density in *Planck* and IRAS channels
	Extraction quality parameters
	[NPSNR, RELTH, SRCSIG, EST_QUALITY]
Source	New location
(free model)	Flux density in *Planck* and IRAS channels
	Thermal SED properties [(a)]
	[Temperature, Spectral index, Ref. Flux Density]
Background	Surface brightness in *Planck* and IRAS channels
(32 × 32 pixel patch)	Signal to noise ratios (source/background)
	Thermal SED properties
	[Temperature, Spectral index, Ref. Flux Density]

Notes. All physical parameters include corresponding uncertainties. [(a)]Fitted to the flux densities *after* extraction.

for example, BeeP cannot find the likelihood peak. In that case there is no guarantee that the recovered parameter estimates are optimal.

6.2. Source properties

This set of parameters gives the position and properties of the sources and their uncertainties.

6.2.1. Thermal properties

We fit the multifrequency data for a given source with two SED models (see Fig. 7), each of which requires an independent run of the likelihood.

- *Modified Blackbody (MBB) model*. The source brightness levels are colour-corrected to account for the detector bandpasses. The following parameters are optimized by the likelihood:
  - X and Y position coordinates, with origin at the PCCS2+2E position;
  - EXT, source extension;
  - SREF, source reference flux density;
  - TEMP, source temperature;
  - BETA, source spectral index.

  All source parameters, geometrical and physical, are sampled jointly. The reference flux density is given at 857 GHz. The reference flux density at 857 GHz is not the flux density measured in the 857 GHz channel; it is rather a scaling factor for the model that could be specified at any frequency. We have chosen 857 GHz for convenience (see Eq. (4)). For this model we also provide the flux densities in the individual channels, computed from the fitted model.
- *Free model*. The FREE columns are developed in two steps. First, samples are drawn from the geometrical parameters and flux densities at each channel. The flux densities at individual channels are optimized by the likelihood. All source parameters, geometrical and physical, are sampled jointly. From the flux-density samples at each frequency we compute a best-fit value and an uncertainty. The following parameters are optimized by the likelihood:
  - X and Y position coordinates, with the origin at the PCCS2+2E position;
  - EXT, source extension;
  - FREES3000, flux density at 3000 GHz;
  - FREES857, flux density at 857 GHz;
  - FREES545, flux density at 545 GHz;
  - FREES353, flux density at 353 GHz.

  We then fit an MBB model to the four data pairs (S_ν, σ_{S_ν}), using a Gaussian likelihood with colour-correction, resulting in a source reference flux density given at 857 GHz.

BeeP also provides, as an output, plots of the source-parameter posterior distributions for the MBB model (see e.g., Fig. 8).

6.2.2. Size

The spatial extent of source-related emission peaks in the maps results from the convolution of the source size and the beam. These are degenerate variables over the relatively narrow range of variation of beam size in the *Planck* maps. BeeP uses a source-extension parameter EXT which represents the intrinsic radius of the source in Eq. (3). However, in Appendix A.2.3, we explain that BeeP artificially narrows the beams to allow for emission bumps in the maps that are narrower than the beam size. Therefore EXT does not correspond to the actual intrinsic source size; however, EXT is easily corrected to a new parameter R, which is the intrinsic source radius corresponding to the real beam sizes. Both parameters are provided in the BeeP results. Furthermore, we remind the reader that we have simplified the source model by assuming that it is a symmetrical 2D Gaussian. The parameter R thus gives a useful indication of whether the source is extended, but it does not reflect any potential source elongation and should therefore be used with appropriate caution.

The distribution of source radii (R) found by BeeP is shown in Fig. 9. The PCCS2 subset (shown in blue), is compatible with a population overwhelmingly dominated by unresolved sources (the size distribution peaks at 1.′2). Instead, the full PCCS2+2E (purple) set peaks at 1.′7. This is expected, since a large fraction of the PCCS2E objects are nearby and Galactic, and many of them show more extended shapes.

6.2.3. Position

One of the important characteristics of BeeP is its ability to determine an effective sub-pixel source position. Since the position is determined from a multifrequency analysis, it does not in general correspond to any of the positions found in PCCS2+2E. POSERR is the uncertainty radius around the position. Its probability density function is a Rayleigh distribution with a scaling parameter equal to POSERR. If $Z = \sqrt{X^2 + Y^2}$ and $\{X, Y\}$ are independent and both normally distributed with a standard deviation σ, then Z follows a Rayleigh distribution with a scaling parameter equal to σ. BeeP's sub-pixel accuracy significantly reduces the large negative kurtosis usually imposed by the pixelization on the error distributions, as can be seen in Fig. 8. POSERR is computed as the 95th percentile of the samples' radial offset distribution divided by 2.45, to give σ, the Rayleigh scale factor. The probability that the true source position is inside a radius of (1×, 2×, 3×) POSERR is (39.3%, 86.5%, 98.9%). Figure 10 shows the dependence of POSERR on NPSNR.

Simulations show that POSERR is significantly underestimated in a subset of cases, predominantly those with high values of NPSNR. A detailed description of this issue is given in Appendix B.2 and shown in Fig. B.6. To address this problem,

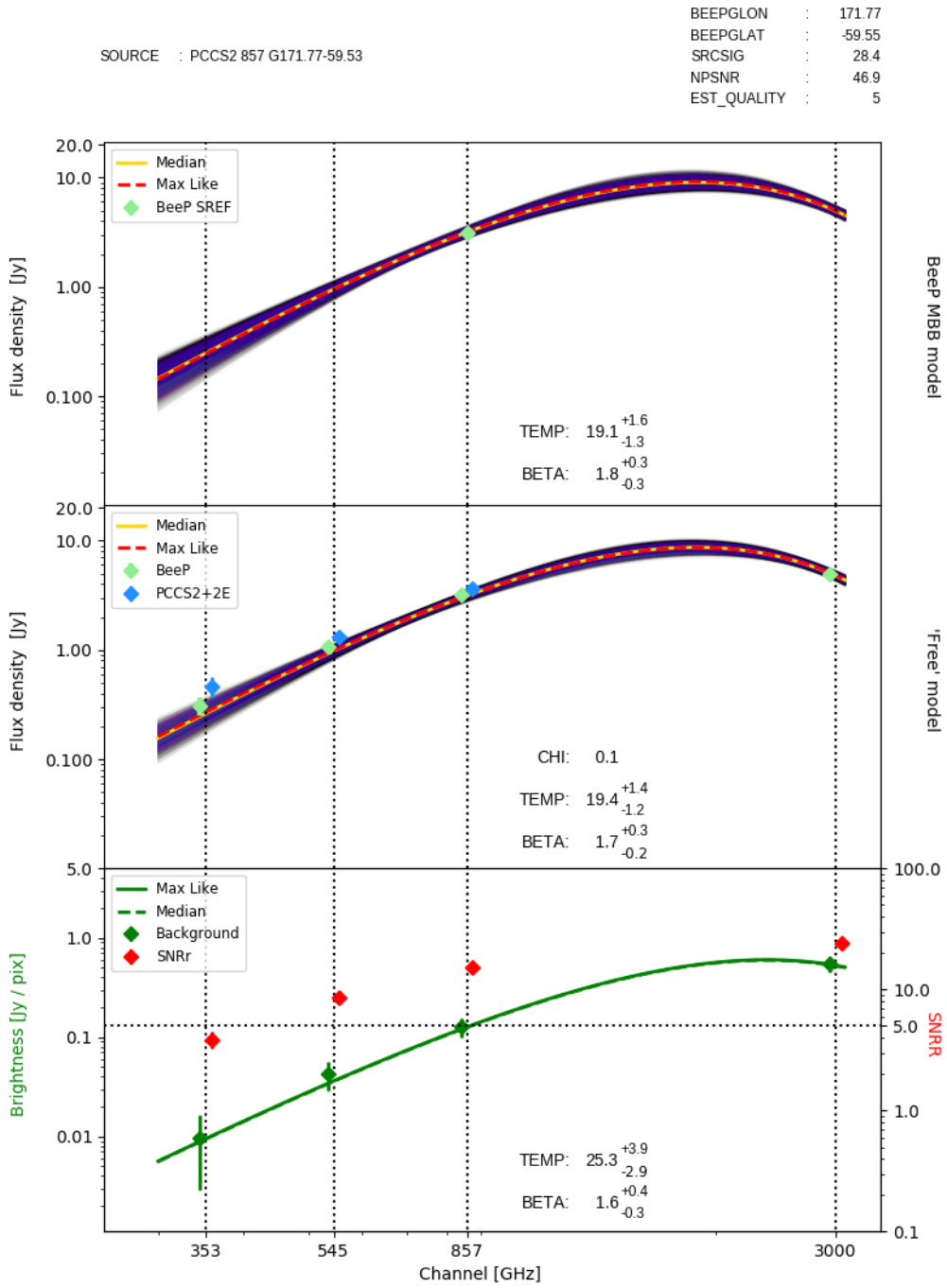

Fig. 7. Example of fitting the MBB (*upper panel*) and Free (*middle panel*) SED models to the data for one source (NGC 895). The background is given in the bottom panel. The yellow and red dashed curves are the median and maximum-likelihood fits, respectively. The purple and black bands are the $\pm 1\,\sigma$ and $\pm 2\,\sigma$ regions, respectively, of the posterior density. Blue diamonds are the PCCS2+2E flux-density estimates (APERFLUX). The green diamonds are: in the *upper panel* BeeP's estimate of the flux density at 857 GHz, and in the *middle panel* BeeP's Free estimates of the flux density at each frequency. In the *lower panel*, dark green diamonds are the background brightness estimates at each frequency, and the green curves are the maximum likelihood (dashed) and the median (solid) models. Red diamonds are the average source brightness divided by the background rms brightness in that patch, i.e., raw S/N. The data points are slightly displaced from their nominal frequencies to avoid overlaps. A similar plot is provided in the *Planck* Legacy Archive for each source in the BeeP catalogue; see the *Planck* Explanatory Supplement for further information (http://www.cosmos.esa.int/web/planck/pla/). We note that this figure is reproduced exactly as it will be delivered to the user from the online archive. In Appendix D we provide some representative examples of spectra for different kinds of sources, to show some of the results obtained by BeeP.

we correct the position errors using the procedure developed in Appendices B.2 and B.4, which follows closely that used for PCCS2 (see Eq. (7) and Table 8 of Planck Collaboration XXVI 2016). The correction consists of adding a term in quadrature to POSERR, which causes small values to saturate at a minimum level of $\sigma_0 = 4''.4$ (see Fig. 10). This level was determined through simulations, as described in Appendices B.2 and B.4.

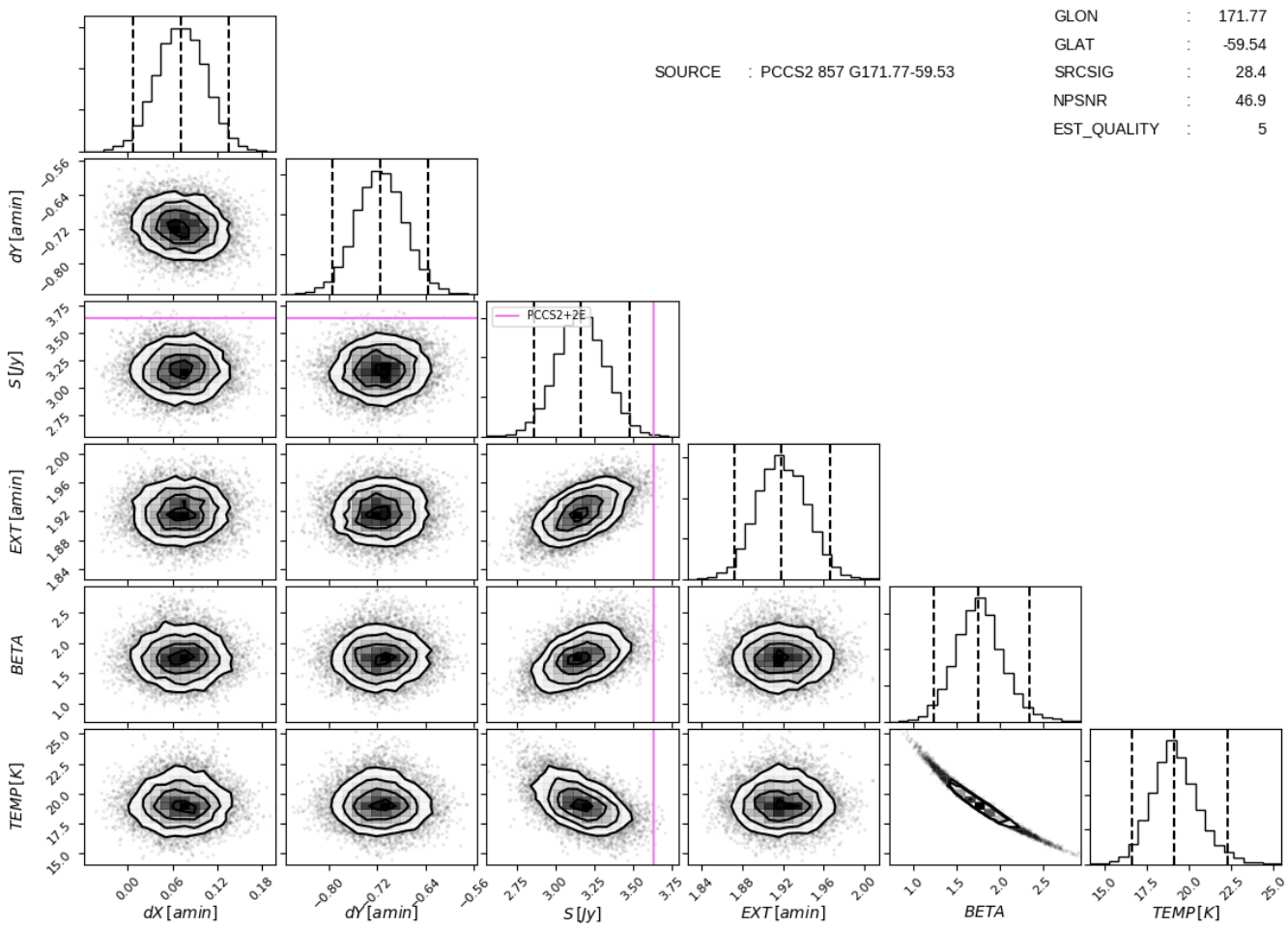

Fig. 8. Corner plot (Foreman-Mackey 2016) of parameter posterior distributions for one source (NGC 895). Off-diagonal positions show marginalized bi-dimensional posterior distributions of the parameter samples defining the row and the column. Diagonal positions contain posterior marginalized distributions. The magenta lines mark the PCCS2+2E catalogue flux density in the 857 GHz channel. There is one such plot for each source in BeeP's catalogue. The source extension (EXT) samples shown have not been corrected for the narrower beams employed in the likelihood. See the *Planck* Explanatory Supplement for further information (http://www.cosmos.esa.int/web/planck/pla/). This figure is reproduced exactly as it will be delivered to the user from the online archive.

To verify that the correction determined through simulations applies to the BeeP/base catalogue, we examined the PCCS2 subset. The correlation seen in Fig. 10 (yellow dots) is very high (−0.98), and its slope $a = -1.09$ is very close to what is seen in the simulations. This high degree of consistency between the simulated data and the real data justifies application of the correction to the data.

The median positional error of the full corrected catalogue is 11.″5 (1/9 of a *Planck* pixel). For the PCCS2 subset it is 7.″9, or less than 1/12 of a pixel.

6.2.4. Flux density

To obtain an unbiased estimate of a flux density, one must know the shape of the instrumental beam and the morphology of the source. By using a constant Gaussian shape to model the beam, equal to the average *Planck* Gaussian effective beam (Mitra et al. 2011), we introduce a systematic bias in estimates of the flux density (see, e.g., Planck Collaboration XXVI 2016, Sect. 2 and Table 2). Furthermore, in any multi-channel analysis such as BeeP, the beam shape is not as clearly defined as in the case of a single-channel catalogue. The effective beam is in fact a combination of the individual channel beams, and it changes with the beam spatial Fourier mode (via the covariance) and source SED parameters. A simple correction such as the one suggested in Planck Collaboration XXVI (2016) is insufficient in this case. Instead, our approach is to "calibrate" the bias in the output of BeeP using simulations. This is explained in detail in Appendix B.4 (see also Appendix A.2.3). The simulations that we use are the *Planck* FFP8 simulations, which are the most complete and realistic for *Planck* 2015 data, and which contain accurate sky and instrument models. Using the FFP8 simulations (Appendix B.4), and comparing recovered values to input values, we estimate that BeeP's reference flux-density estimator is biased high by about 11.0%, which reflects the lack of realism of our model regarding source extension. An 11% reduction in the reference flux densities produced by BeeP is therefore applied to both SED models (MBB and Free). Specifically, flux densities in all four channels are reduced by this same factor for the Free model.

The estimated flux-density accuracy is also subject to systematic effects caused by beam and source shapes. Figure 11 displays the variation of the relative flux-density error bar $\sigma_S^{\rm rel}$, defined as

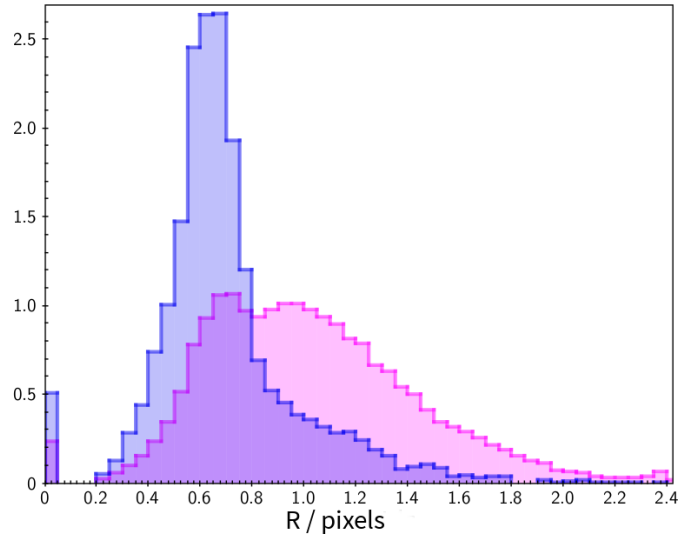

Fig. 9. Normalized histograms of the recovered source size R PCCS2 sources are shown in blue and the full catalogue in purple. R has been corrected for the excess resulting from using narrower beams in the likelihood. Beam-sized objects appear in the figure at R ∼ 0. One pixel here corresponds to 1.′72.

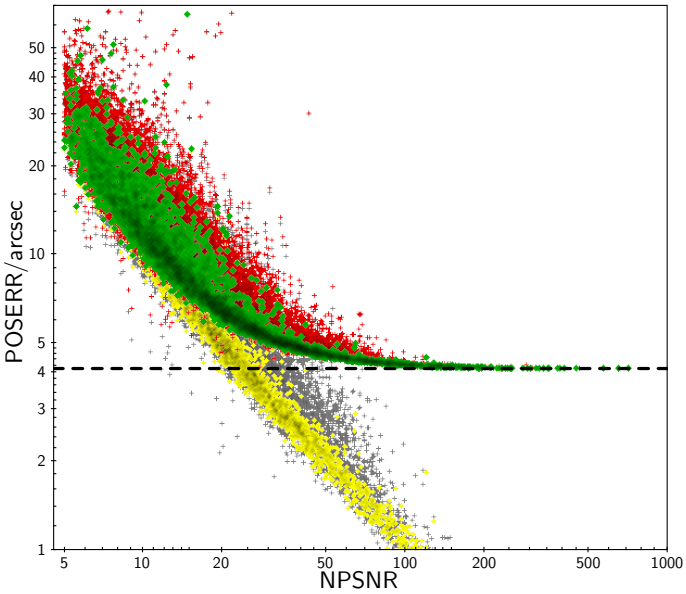

Fig. 10. Radial position error POSERR versus NPSNR. Grey and yellow points mark sources in the PCCS2+2E and PCCS2, respectively, before correction. Red and green points mark sources in the PCCS2+2E and PCCS2, respectively, after correction. The horizontal dashed line is the saturation constant added to correct the position uncertainty, 4.″11.

$$\sigma_S^{\rm rel} \equiv \frac{\Delta S}{S}\,, \tag{17}$$

where S is the estimated flux density, ΔS is the estimated flux-density uncertainty, and $\sigma_S^{\rm rel}$ is the inverse of the measured S/N. For reference, the black dashed line on the left lower corner is the NPSNR^{-1} line. This is the theoretical lower boundary for $\sigma_S^{\rm rel}$ that would be expected if the only unknown parameter were the flux density. Figure 11 shows that the catalogue's flux-density uncertainties are much higher ($\sigma_S^{\rm rel} \gg \mathsf{NPSNR}^{-1}$) than the lower boundary, which should be expected from the fact that there are five more unknown parameters, whose individual uncertainties propagate into the flux-density estimate. However, not all of the additional parameters contribute equally. Inspecting the posteriors in Fig. 8, it becomes clear that EXT and the MBB parameters $\{T, \beta\}$ have a much larger contribution than the position parameters. The correlation between the flux errors and the other parameter uncertainties explains the gap between the black dashed line and the green points in the figure. However, with the help of simulations (see Appendices B.3 and B.4), we find that the estimated flux-density errors are overly optimistic for a fraction of the high NPSNR population. The situation is similar to that for the positional accuracy estimates (see Sect. 6.2.3). For most purposes the (uncorrected) flux-density estimates and uncertainties found in the catalogue can be used without concern. But if a more rigorous statistical characterization is required, we suggest correcting the flux-density uncertainty estimates using the procedure developed in Appendix B.3. There is a modest penalty in flux-density accuracy for applying this correction (Fig. 11, red contours).

BeeP produces two sets of flux-density estimates: the MBB SREF and the Free FREESREF. In Fig. 12 we compare their values to test the consistency between the two models. Instead of simply calculating percentage differences, we plot the logarithm of the output to input ratio. If out/in ≈ 1, then ln(out/in) ∼ (out − in)/in, which corresponds closely to percentages. But when out/in is far from 1, then ln(out/in) keeps the symmetry between in and out, which would not be the case with the more common (out − in)/in formula. We find this feature very convenient for visually identifying biases in the differences.

As expected, there is higher dispersion for sources drawn from the PCCS2E catalogue (shown in red), as a result of generally more complex backgrounds at low Galactic latitudes. Sources from the PCCS2 (shown in green) are less affected by this issue. We note the small (3.5%) bias towards negative values of $\ln(S_{\rm Free}/S_{\rm MBB})$. This bias becomes more pronounced at lower values of SRCSIG. A possible source of this bias is that inclusion of inter-frequency cross-correlations in the likelihood for the background model allows for better removal of background emission, on average raising $S_{\rm MBB}$.

6.2.5. Spatial distribution of the source properties

Figure 13 shows the spatial distribution on the sphere of MBB prameters T and β for the compact sources. High Galactic latitudes show a larger percentage of warmer objects (see Fig. 14) and very few cold sources (dark blue). Cold sources are mostly Galactic in nature, and aligned with filaments of gas and dust. They also match regions of intense star formation. As a result of the strong correlation between T and β (see Sect. 6.2.6 and Fig. 8), a higher density of sources with low β is expected at higher Galactic latitudes.

One problem with the extraction of this catalogue is the severe non-homogeneity of the background. The brighter sources, represented with larger circles, are concentrated in the Galactic plane (see Fig. 13). However, as one can see in Fig. 15, the regions with higher SRCSIG are preferentially located at high Galactic latitudes, roughly matching the PCCS2 domains. This is the result of smoother backgrounds and less severe non-Gaussianity.

6.2.6. Source populations

Figure 16 (left panel) shows the catalogue MBB estimates on the T–β plane, coloured by Galactic latitude. The T–β set forms a banana-shaped distribution with an excess of colder sources

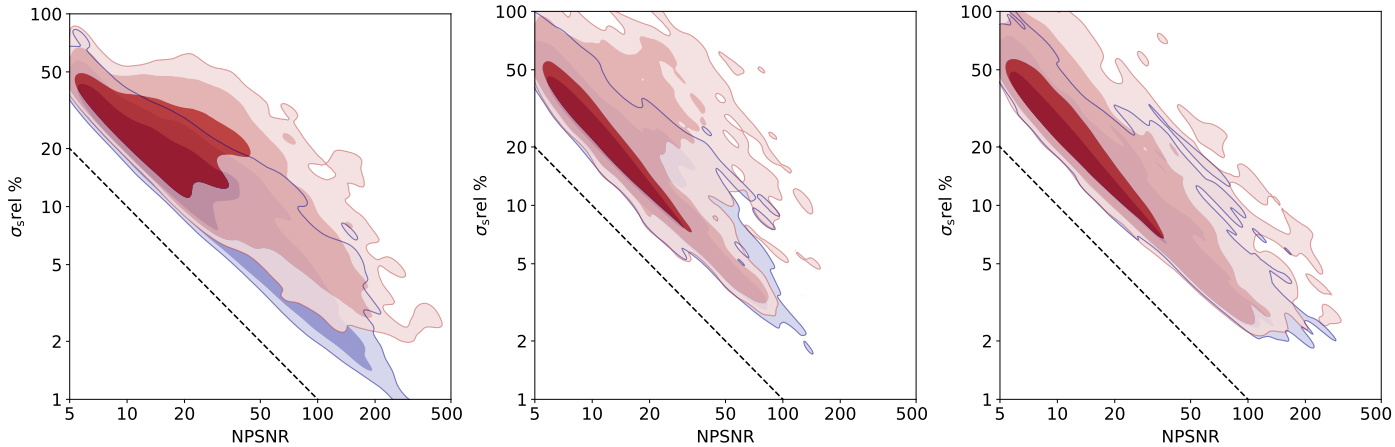

Fig. 11. Flux density uncertainties ($\sigma_S^{\rm rel}$, see Eq. (17), as a percentage) versus NPSNR. Blue contours ([68,95,99]%) show the distribution of uncorrected values (as presented in the catalogue), while red contours show the distribution of values after the corrections suggested in Appendix B.2. *Left panel*: sources in the PCCS2, *middle panel*: PCCS2E with Galactic latitude greater than 10°, and *right panel*: sources close to the Galactic plane, with $b \leq 10°$. For reference, we show $\mathtt{NPSNR}^{-1}$ (black dashed line), the theoretical lower boundary for $\sigma_S^{\rm rel}$, which can only be achieved if flux density is the sole parameter in the model.

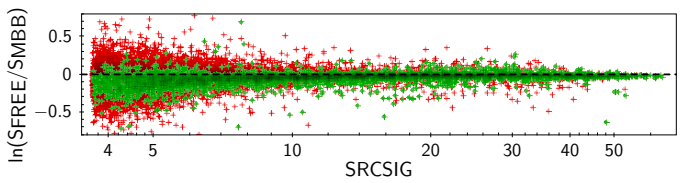

Fig. 12. Ratio of free flux density to MBB model flux density at 857 GHz, as a function of SRCSIG. PCCS2 sources are shown in green, while PCCS2+2E are in red. Only sources whose Free and MBB positions are within 0.′8 (half a pixel) of each other, and whose MBB fit to the independent flux measurements has reduced $\chi^2 < 5$ are included. The total number of sources included (25 236) are ≈ 97% of the `BeeP/base` catalogue.

($T < 18$ K) at low Galactic latitudes. This cold population was the main target of the *Planck* Catalogue of Galactic Cold Clumps (GCC, Planck Collaboration XXVIII 2016, see Sect. 7). `BeeP`'s likelihood has a more inclusive selection criterion, since it is not limited to sources embedded in warmer backgrounds. However, as may be seen in Sect. 7, the temperature contrast between source and background boosts the detection strength (NPSNR). The T–β uncertainty (in grey) can be important, particularly for warmer ($T \geq 18$ K) and steeper sources ($\beta \geq 1.5$). Most of the warmer sources are faint at 353 and 545 GHz, such that they are just above or even below the background levels. This severely reduces `BeeP`'s constraining power, since then only the two higher frequency channels contribute significantly.

Figure 16 (right panel) illustrates more clearly the influence of individual channels on the overall significance, which is affected by all the channels processed by `BeeP`. Colder sources are brighter, for the same reference flux density, at the three *Planck* channels, which have much lower noise than IRIS. Naively, one would thus expect these sources to have high reliability; however, they are also much more likely to be found embedded in bright and complex background regions at low Galactic latitude. This imposes upon them a penalty on source significance, not only because the background is stronger, but because the levels of non-Gaussianity are also much higher. For this reason, colder sources have generally lower estimated SRCSIG than warmer ones.

There is a small group of synchrotron flat-spectrum sources characterized by their non-physical MBB parameter values (see Fig. 16, bottom left corner of both panels). To identify this subset, we found all sources that satisfied BETA < 0.5, TEMP < 15, and EST_QUALITY ≥ 4. We cross-matched the high-Galactic-latitude ($|b| > 20°$), flat-spectrum population (24 sources) with *Planck*'s PCCS2+2E 30 GHz catalogue. The cross-match returned 23 common objects. Note that we are not removing any of these sources, but providing a simple way to identify them in the extended catalogue. The remaining `BeeP` object (PCCS2 857 G207.16-60.71) just misses the reliability criterion of `BeeP/base`, with NPSNR = 4.84 < 5.

6.3. Background properties

As a by-product of the `BeeP` analysis, we obtain the MBB parameters of the background thermal emission around each source. We compute the average brightness and standard deviation $\{I_\nu, \sigma_{I_\nu}\}$ from the four background maps over a square patch 33 × 33 pixels (56.′7 × 56.′7) across, centred on the PCCS2+2E source position. Reduced resolution was also employed in (Planck Collaboration XI 2014) to stabilize the evaluation of the T–β pairs. The CIB monopole, added to the *Planck* 2015 maps as reported in Planck Collaboration VIII (2016), was then subtracted. Offsets do not affect estimates of properties of compact objects, as they are subtracted before the likelihood evaluation. However, they are important when estimating the background thermal properties. Uncertainties resulting from map calibration and CIB monopole errors are also added directly to σ_{I_ν}. Then, following exactly the same procedure as in the case of the Free source model, an MBB background model curve, with colour correction, was fitted to these data pairs using a Gaussian likelihood. These curves are also shown in the SED plots, e.g., Fig. 7.

At these frequencies the dominant background component is dust, particularly for low Galactic latitudes (Fig. 17 left and centre). However the picture becomes slightly more complicated for high Galactic latitudes where CIB anisotropies, instrumental noise, and CMB anisotropies (especially at 353 GHz) also make significant contributions. CIB anisotropies are important only at scales of 1° and smaller (Planck Collaboration Int. XLVIII 2016; Planck Collaboration XI 2014), while instrumental noise

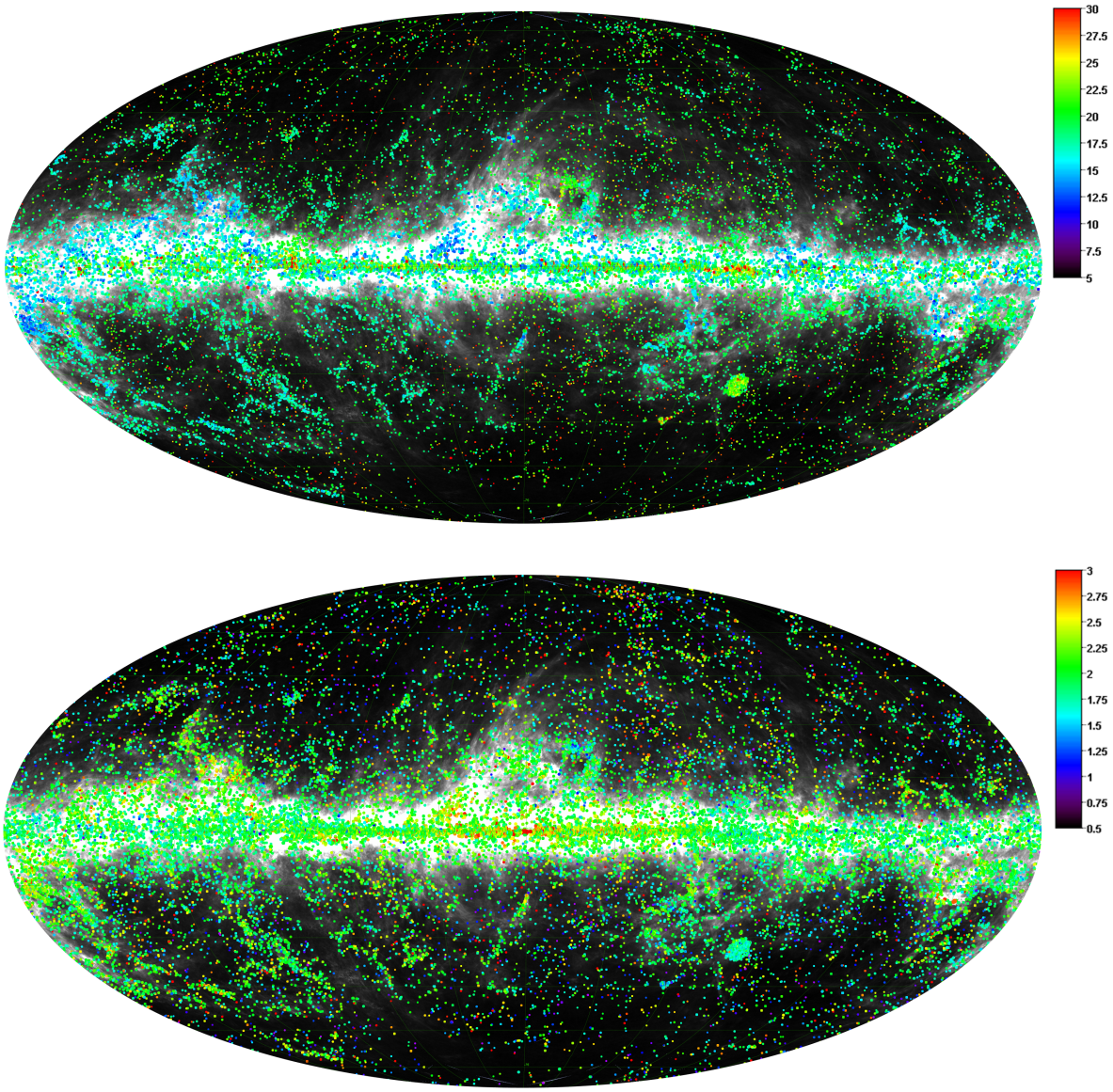

Fig. 13. *Top*: temperatures of sources in the catalogue (colour scale in thermodynamic kelvins). *Bottom*: spectral indices β of the MBB SED model. The catalogue was filtered using the condition of Eq. (16). The size of each circle representing an object is proportional to the logarithm of the source flux density in janskys. This figure also makes clear the extent that sources in PCCS2+2E trace cirrus; see also the smaller region in Fig. 1.

is important only at even smaller scales. In contrast, the CMB appears predominantly at 1° and larger scales. The CMB signal is faint compared to dust at 545 GHz and above, but not at 353 GHz at high Galactic latitude.

Figure 18 shows the histogram of the MBB parameters T and β for dust-rich regions defined by the masks used in PCCS2E (blue), and for high Galactic latitudes (green). The distributions are different as a result of their dissimilar composition (see also Fig. 17 left). In regions where dust is dominant, the agreement with Generalized Needlet Internal Linear Combination (GNILC) (Planck Collaboration Int. XLVIII 2016) estimates of temperature and β is excellent. In regions of low column density, the agreement deteriorates significantly (see also Sect. 7.4 and Fig. 29 for further explanation of these features).

Figure 19 shows the spatial distribution of the estimated background parameters. High-Galactic-latitude zones show consistently higher temperature and lower β than the dusty regions close to the Galactic plane. This result is consistent with previous analyses (Planck Collaboration XI 2014; Planck Collaboration XIX 2011). Regions close to the Galactic centre have higher β than those at larger longitudes (upper right panel). The effect is less pronounced for T (upper left panel).

One of the interesting background parameters estimated by BeeP is RELTH, which measures the non-Gaussianity of the background. The spatial distribution of RELTH is shown in the bottom right panel of Fig. 19. As a consequence of the non-Gaussian nature of dust emission, it is expected that RELTH correlates with background emission, and this is indeed evident from the bottom panels. Nonetheless, RELTH depends on the detailed statistics of the field being analysed (Sect. 3.3, Eq. (9)), therefore direct comparison of RELTH levels in regions of widely varying complexity is likely biased. Figure 17 (right

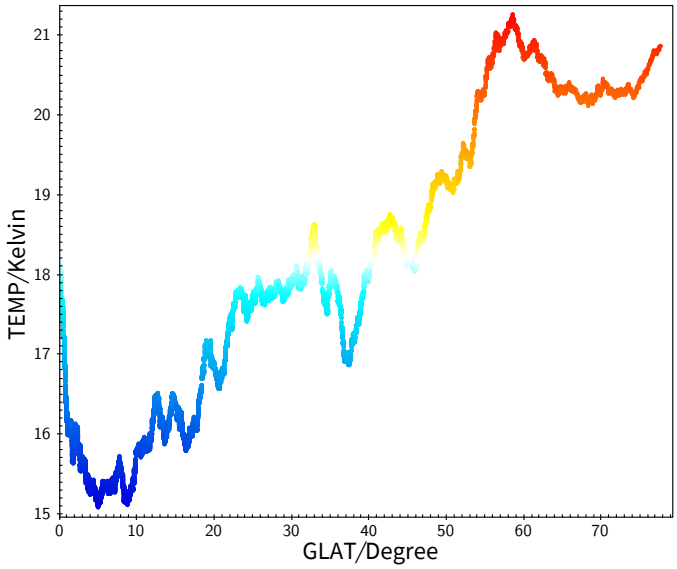

Fig. 14. Boxcar average (window 500 samples) temperature of sources ordered by absolute Galactic latitude. There is a clear trend.

panel) shows that although regions with high non-Gaussianity exist over the full range of thermal emission properties, the coldest background regions are *all* highly non-Gaussian. This is at least partly due to the fact that they are located near to the Galactic plane, where there is the most confusion.

7. Comparison with other catalogues

7.1. Planck PCCS2 catalogue

The PCCS2 contains the most reliable sources in the full PCCS2+2E, because of their location in low-background regions. The PCCS2+2E was built using a single channel MHW algorithm, which is of a very different nature than that of BeeP. Therefore, comparison of PCCS2 and BeeP source parameters provides an interesting cross-validation of the two methods. For reference, we also include PCCS2E in the comparisons.

7.1.1. Flux density estimates

To compare source flux-density estimates, we use the BeeP Free values at all but 857 GHz, since these were obtained with a data model more in line with the single-channel measurements of the PCCS2+2E. At 857 GHz, however, we compare the fitted MBB flux density, not the individual flux density (FREES857). This allows for a broader validation because we are testing the full range of flux densities and the SED model all at once. It is also a less noisy estimate. At the same time, since so much more information goes into the BeeP estimate than into the PCCS2 estimate, we should not a priori expect a very good match. For PCCS2+2E, we use APERFLUX estimates, which were obtained using an aperture photometry algorithm.

Figure 20 shows the results of this comparison. On the average, there is good consistency between BeeP's estimates and those in the PCCS2. There is, however, an overall bias with a median of about +4.0% (mean +5.8%) for PCCS2, which increases to +5.0% (mean +7.3%) for the full PCCS2+2E.

Although we use the PCCS2+2E source locations as a starting point, we allow BeeP to search for a better effective location in the close neighbourhood if that increases the likelihood ratio. Maximizing the likelihood ratio is equivalent to maximizing the flux density, because we assume that the background is homogeneous around the source. One might expect that the population of sources that moves from its original position by a significant amount should, on average, show higher flux densities. Figure 20 shows in red sources that moved by more than one pixel from their PCCS2+2E position. As expected, the densities of this population are clearly biased high compared with those of PCCS2+2E. Considering these effects, we remove from the comparison sources whose position changed by more than one pixel with regard to the PCCS2+2E estimate, and extended sources with R $\geq$ 1.64 pixels. Values of R in Fig. 9 were obtained from EXT, correcting for the excess that results from using narrower beams in the likelihood. Now the flux-density bias for the full PCCS2+2E becomes negligible: median = −0.6% (mean = −0.8%). Therefore from now on we only use this subset of sources to compare BeeP's flux-density estimates with those in the PCCS2+2E catalogue.

The second factor affecting the flux bias between BeeP and PCCS2 is background removal. For low APERFLUX, BeeP's flux densities seem to become increasingly biased high as we go down in flux. At 0.45 Jy we are already at the sensitivity limit for single-channel aperture photometry. At these very low flux densities, the effects of Eddington-type bias become important. However, the multi-channel nature of BeeP makes it more sensitive, with an efficient background removal even at these low signal regimes. This effect is much more pronounced in the PCCS2 than in the PCCS2E, because the fraction of sources with flux densities below the 0.45 Jy threshold is larger. A simple example to understand how this bias occurs is the following. Imagine a completely homogeneous but positively correlated background, containing valleys and crests. Now imagine a very faint source population, all of the same flux, embedded in it. Applying an aperture photometry method to recover the flux, the sources sitting in the valleys, would appear in the faint end group. A method that could reduce the background to zero would recover the true flux, which when compared with the faint end of the aperture photometry flux estimate would appear biased high. To account for that, in addition to the previous filters, we further removed sources with APERFLUX < 0.45 Jy. The comparison restricted to this PCCS2 subset now shows a rather small bias: median = 1.1% (mean = −1.1%).

On average, the uncertainties in BeeP flux densities are a factor of about 2 smaller than those of the aperture flux estimates in PCCS2+2E (see Fig. 21). The combination of uncertainties obtained by BeeP and those of PCCS2+2E explains the dispersion of Fig. 20 adequately.

We further compared BeeP's Free flux-density estimates at 353 and 545 GHz, FREES353 and FREES545, with the PCCS2+2E equivalents, APERFLUX_353 and APERFLUX_545 (see Fig. 22). The subset depicted was obtained by removing sources whose BeeP position estimate changed by more than one pixel from the original PCCS2+2E and those that appear to be extended, with R $\geq$ 1.64 pixels (see Fig. 9). For the PCCS2+2E (in purple) the flux-density biases we find are 0.6% (median) at 545 GHz and −2.6% (median) at 353 GHz. The dispersion of the estimates is high, similar to that of the 857 GHz channel, but consistent with the combined uncertainties from BeeP and PCCS2+2E.

7.1.2. Source positions

Assuming that positional errors $\{X, Y\}$ are independent and Gaussian-distributed in both the PCCS2+2E and BeeP, then the

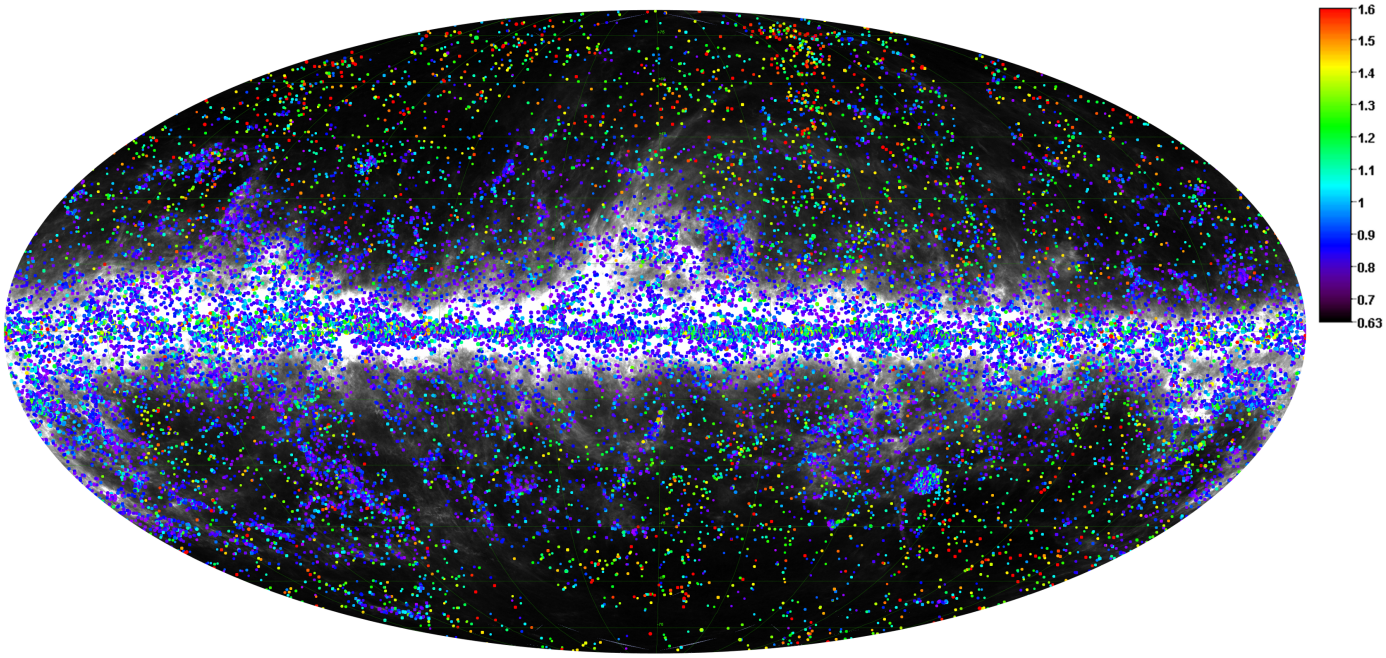

Fig. 15. Spatial distribution of the significance statistic SRCSIG. The colour bar represents SRCSIG on a logarithmic scale.

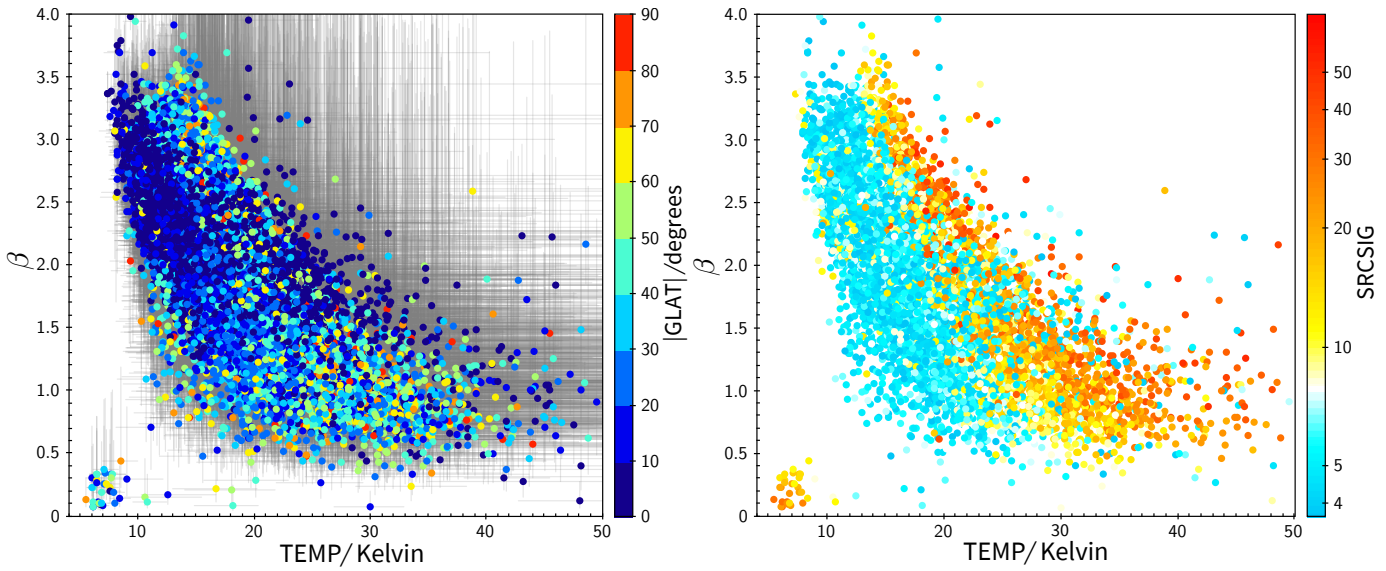

Fig. 16. MBB parameters β and T for sources in the catalogue. The sources are coloured by Galactic latitude in the *left panel*, with 1 σ error bars in grey, and by the SRCSIG statistic in the *right panel*. The small cluster of sources close to the lower left corner is the non-thermal population of flat-spectrum sources.

distance between both positions should follow a Rayleigh distribution (see Sect. 6.2.3) with a scale factor σ dependent on the positional accuracies of both catalogues. Figure 23 shows the histogram of the distances between the BeeP and PCCS2+2E positions. The PCCS2 subset histogram (in blue) is a good match with the shape of a Rayleigh distribution. As expected, the PCCS2E exhibits a wider tail. The PCCS2E histogram also has bumps at 1.′72 (1 pixel) and 3.′43 (2 pixels). These small excesses are the natural result of the map pixel grid. As may be seen in Fig. 8, BeeP's positional uncertainty seems little affected by the map pixelization. However, the presence of these small bumps at exact multiples of the pixel size, indicates a possible greater impact on the PCCS2+2E, which might add a small negative kurtosis in the PCCS2+2E $\{X, Y\}$ error distributions. Given that BeeP's positional uncertainty is so small, if we take the BeeP positions as the true values, then the histograms in Fig. 23 are consistent with the positional uncertainty characterization of the 857 GHz channel in the PCCS2+2E. The distributions (PCCS2 and PCCS2E) peak at around 0.′65. This value is a good match to the average 0.′65 position error estimate for the 857 GHz channel of the PCCS2 subset in Eq. (7) and Table 8 of Planck Collaboration XXVI (2016).

7.1.3. Background complexity and reliability

The PCCS2+2E catalogue contains a field CIRRUS_N that flags entries with a complex background, and therefore a higher probability of being spurious. CIRRUS_N is the number of sources

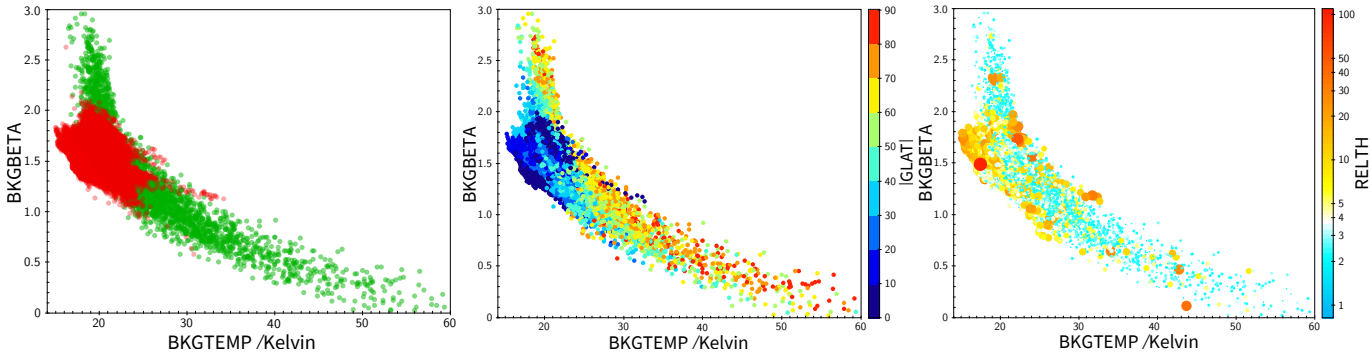

Fig. 17. Three views of the relationship between T and β for the background, from the MBB model. *Left*: PCCS2 background regions, which are at high Galactic latitude, are in green. PCCS2E Galactic mask regions, which have strong, dominant dust emission, are in red. *Middle*: colours show Galactic latitude. *Right*: colours show RELTH, a measure of the non-Gaussianity of the patch. The symbol size here is proportional to RELTH.

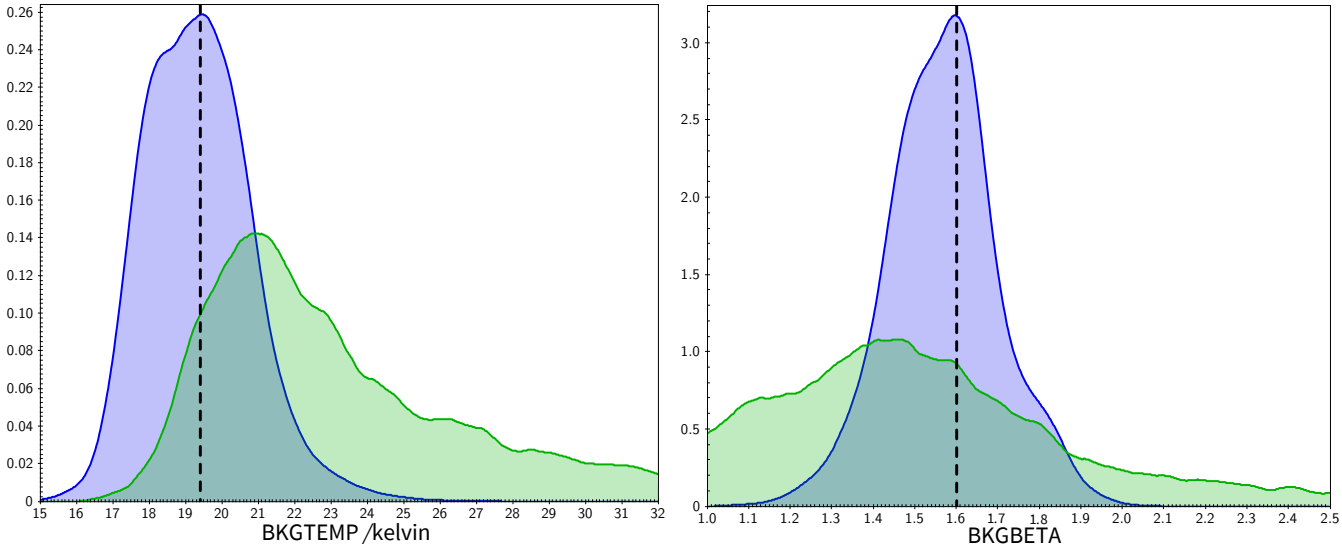

Fig. 18. Distributions of MBB background temperature (*left*) and spectral index (*right*). PCCS2 regions (high Galactic latitude) are green and zones where dust emission is dominant are blue, as defined by the PCCS2E cirrus and Galactic masks. Dashed lines mark the best-fit values for dust temperature (19.4 K) and β (1.6), from Planck Collaboration Int. XLVIII (2016).

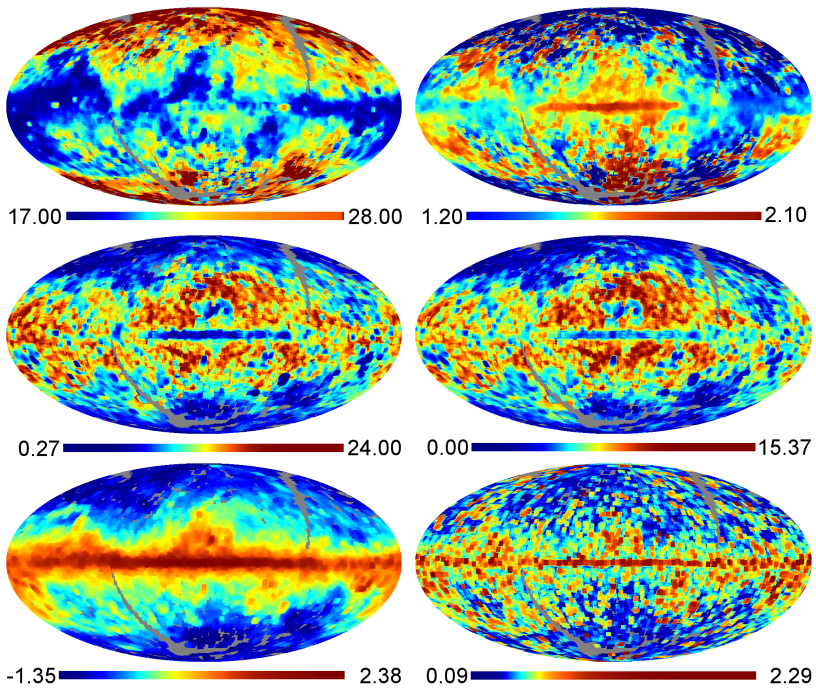

Fig. 19. *Top*: MBB parameters fitted to the background, with T(K) on the *left* and β on the *right*. The high-temperature region at ($l = 155°; b = 77°$) is artificially created by strong artefacts in the IRIS data. *Middle*: inverse relative uncertainty $T/\Delta T$ (*left*) and $\beta/\Delta\beta$ (*right*) of the MBB parameters. *Bottom left*: reference background brightness, log(Jy pixel^{-1}, evaluated at 857 GHz. This is the value at 857 GHz of a multifrequency fit, not the value directly measured at 857 GHz. *Bottom right*: log(RELTH), computed with $\alpha = 5\%$ (see Eq. (9)). On all panels there is a region at ($l = 208°; b = -18°$) of extreme values and uncertainties caused by artefacts present in the IRIS data. The colour bars have been histogram-equalized. regions are either inside the IRIS mask, or had insufficient data.

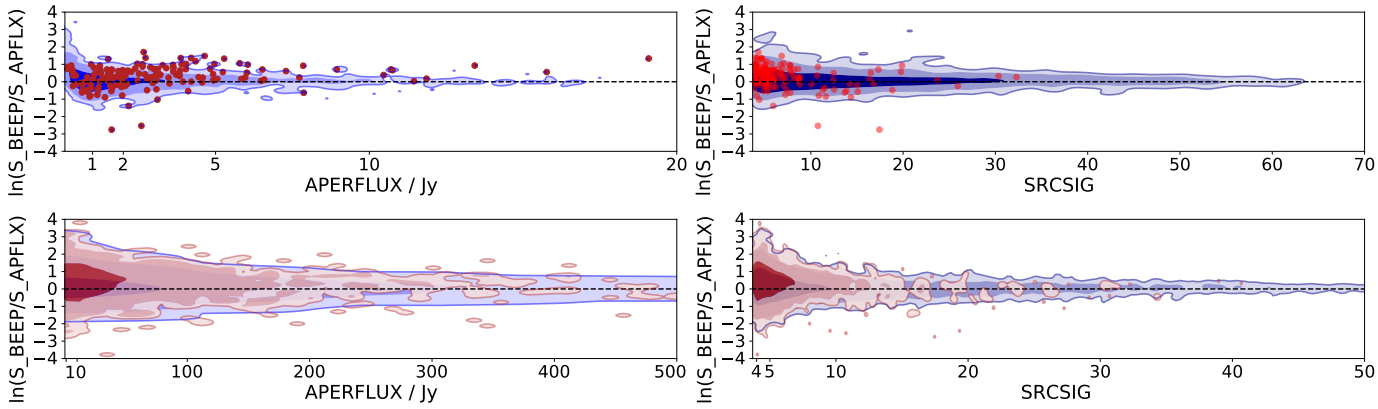

Fig. 20. Comparison of BeeP's Free flux densities at 857 GHz with aperture flux-density values (APERFLUX) from PCCS2+2E. We plot $\ln(S_{\mathrm{BeeP}}/S_{\mathrm{APERFLUX}})$ against APERFLUX (*left*) and SRCSIG (*right*). *Upper row*: distribution of the PCCS2 values in blue (contours are [68, 95, 99]%). The red dots are sources that moved by more than one pixel from the original PCCS2+2E position. *Lower row*: full PCCS2+2E. The red contours represent the distribution of sources that moved by more than one pixel, and the blue ones the remaining population. *Top row*: we show individual dots because there are too few of them to make a density plot.

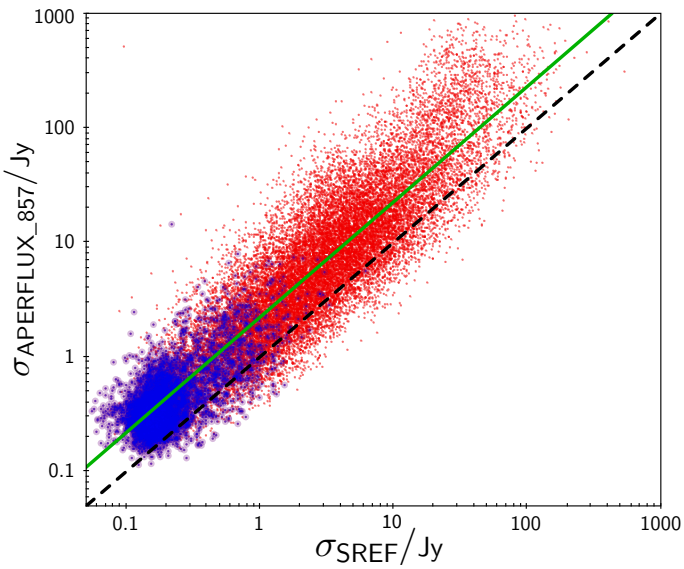

Fig. 21. Comparison of the uncertainty determined in PCCS2+2E (red markers), and the PCCS2 (blue markers), on the aperture flux density at 857 GHz (APERFLUX_857) to the uncertainty on the reference flux density as obtained by BeeP (SREF). The black dashed line represents equality, and the green solid line is the best fit, which has a slope very close to 2 (actually 2.2).

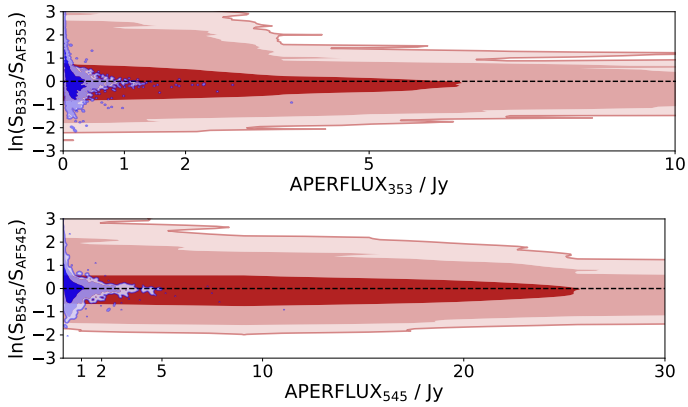

Fig. 22. Comparison of BeeP's Free flux densities at 353 GHz (*upper row*) and 545 GHz (*lower row*) with aperture flux-density values (APERFLUX) from PCCS2+2E. The blue contours ([68, 95, 99]%) show the distribution of the PCCS2 subset of sources and the red ones the remaining PCCS2E. Unlike in Fig. 20, BeeP sources whose position shifted by more than one pixel from the original PCCS2+2E and those with EXT ≥ 1.64 pixels (extended) are not included.

detected at 857 GHz within a circle centred on the source with a radius of 1° (Planck Collaboration XXVI 2016). BeeP's RELTH is a measurement of the local background non-Gaussianity, either intrinsic or as a result of localized structures (cirrus and filaments). Its role is pivotal in defining BeeP's reliability criterion SRCSIG (see Eq. (12)): a higher value of RELTH implies a larger correction to NPSNR, or, similarly to CIRRUS_N, a lower reliability of a putative source. These two quantities, although different, should exhibit some degree of correlation if the background non-Gaussianity is indeed the main source of false positives. Figure 24 shows such a correlation between the variables. The relationship is particularly tight for low values of both variables, as seen in the inset part of the figure, which was obtained by applying the same procedure as in the main picture to the PCCS2 subset. The moving-average window was also reduced to 50 samples for greater resolution. It is clear from this figure that below CIRRUS_N = 8, there is a well defined correlation between the two quantities. The opposite happens above the threshold. According to Planck Collaboration XXVI (2016), CIRRUS_N = 8 is the suggested source reliability threshold.

7.1.4. Reliability

PCCS2+2E contains no reliability information at 857 GHz. The PCCS2 list at 353 GHz is the closest in frequency to BeeP's reference channel (857 GHz), which includes source reliability information. PCCS2 sources at 353 GHz are classified as having medium (80%) to high (99%) reliability. We cross-matched (within a 5′ radius) PCCS2 353 GHz sources with the full BeeP catalogue, and found 786 (58.5% of the PCCS2 353 GHz list)[21].

[21] Most of the unmatched 353 GHz sources do not have a counterpart in the PCCS2 at 857 GHz, due to the important increase in the level and complexity of the background, which reduces the S/N and results in no detection.

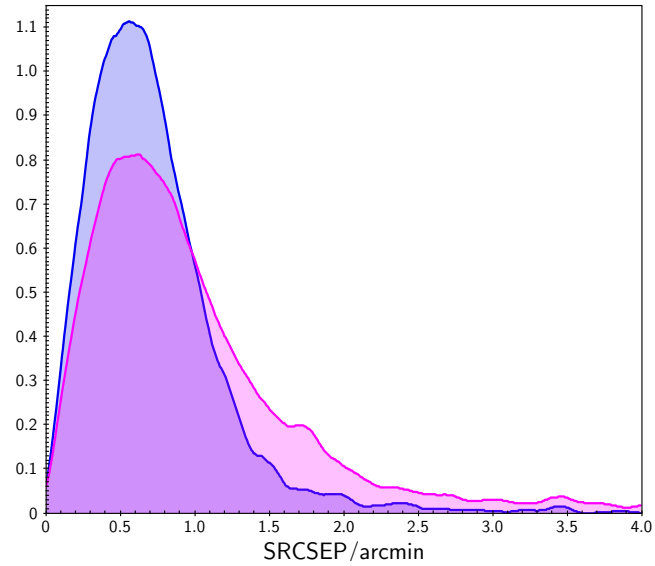

Fig. 23. Normalized histograms of the differences between the PCCS2 (in blue) and the PCCS2E (in pink) position estimates and those from BeeP. The PCCS2E distribution shows a significantly more extended tail. The two small bumps in the distribution are at 1′.72 (1 pixel) and 3′.43 (2 pixels). The PCCS2 histogram has a less extended tail compatible with smoother homogeneous backgrounds. The peak of both distributions is at about 0′.65, the average position-error estimate for the 857 GHz channel in Eq. (7) and Table 8 of Planck Collaboration XXVI (2016).

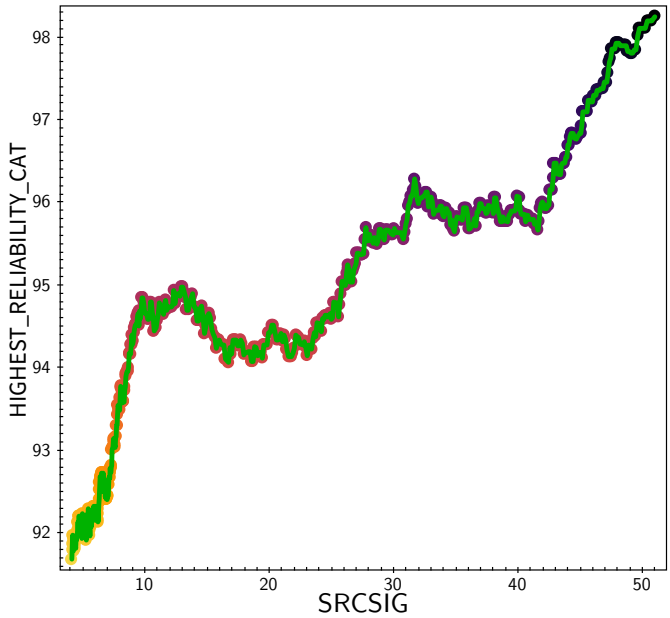

Fig. 25. Correlation between the PCCS2 quantities HIGHEST_RELIABILITY_CAT and BeeP SRSIG, for common sources, as described in the text. We sorted the sources in SRCSIG ascending order, and computed a boxcar average over SRCSIG and HIGHEST_RELIABILITY_CAT with a 100-sample window. There is a clear positive correlation between these two variables.

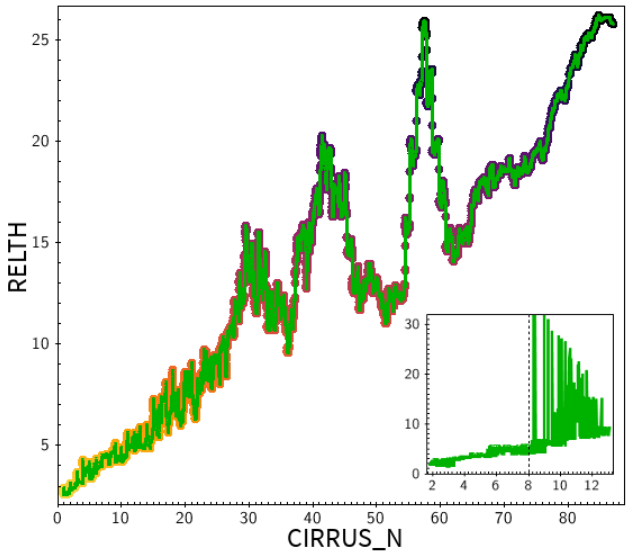

Fig. 24. Correlation between the BeeP RELTH parameter and the PCCS2+2E CIRRUS_N parameter. Sources in the BeeP catalogue were sorted in ascending order of RELTH. Then a boxcar average with window = 500 was calculated for both RELTH and the corresponding PCCS2+2E CIRRUS_N values. The relationship is particularly tight for low values of the parameters, as seen in the expanded detail window, obtained following the exact same procedure as the main picture but using PCCS2 data only. We also reduced the boxcar window to just 50 samples here. The dashed vertical line is the PCCS2 CIRRUS_N reliability threshold.

All but one of these 786 sources appear in the BeeP catalogue with a contamination lower than 10% (SRCSIG > 2 and NPSNR > 5). Demanding an even lower contamination of 5% in the BeeP selection (SRCSIG > 4.0, NPSNR > 5), there remain 744 common sources (94.7% of the common 786 sources). The highest reliability (99%) subset of PCCS2 at 353 GHz contains 427 sources, 416 (97.4% of the PCCS2 353 GHz sources) of which are in the BeeP 5% contamination subset. Given that the majority of the PCCS2 353 GHz catalogue sources have positive spectral indexes (Planck Collaboration XXVI 2016), if they are already reliable at 353 GHz they should be even more so at higher frequencies, where they are brighter. However, one must also consider the effect of the embedding background. If the spectral index of the background is steeper than that of the source, the contrast between the source brightness and that of the background might actually decrease. As a check, we cross-correlated PCCS2 353 GHz HIGHEST_RELIABILITY_CAT with BeeP/base's SRCSIG (Fig. 25). The well defined trend confirms the expected positive correlation.

7.2. Planck Catalogue of Galactic Cold Clumps (GCC).

The *Planck* Catalogue of Galactic Cold Clumps (GCC) was constructed from the same input data used by BeeP, namely, the *Planck* 353, 545, and 857 GHz channels, and the 3000 GHz IRIS map (Planck Collaboration XXVIII 2016). Similarly to BeeP, GCC is generated using a multi-channel algorithm (CoCoCoDeT) on the entire sky. However, this detection algorithm is very different than that of BeeP, since it targets the temperature contrast between cold clumps (cold compact emission regions) and a warm background (Montier et al. 2010). The difference in approach makes it interesting to compare the parameters estimated by BeeP and GCC.

For this purpose, we cross-matched BeeP and the GCC catalogues using a 5′ matching radius. The common set contains 8690 entries (65.6% of GCC). Of these, only 47 are in the PCCS2 (0.54%). If we further require that the common

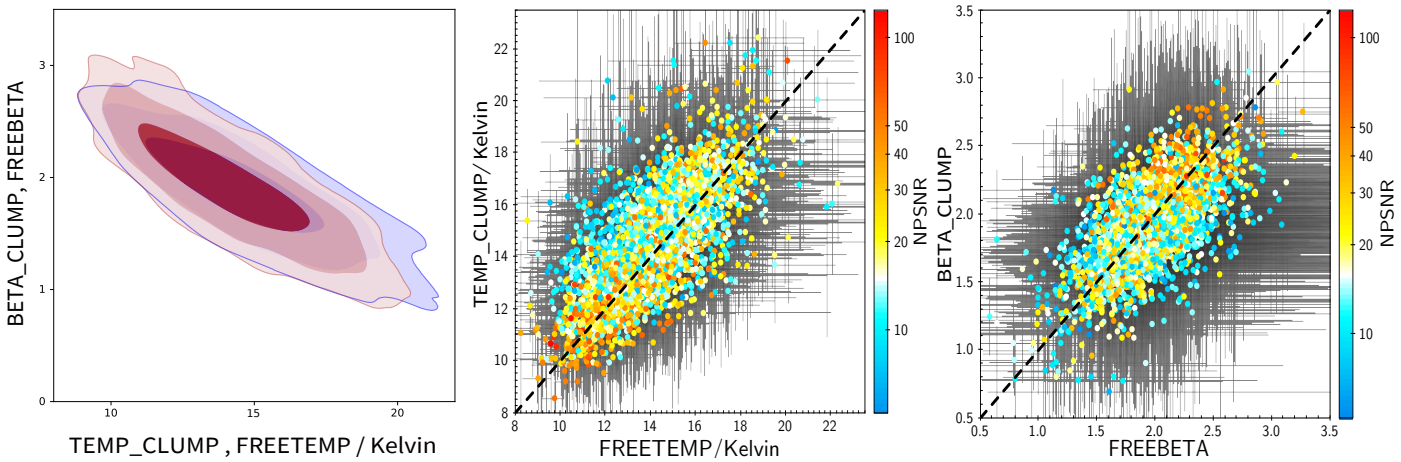

Fig. 26. Comparison of source properties as found by BeeP and in the Galactic Cold Clumps catalogues, for the cross-matched subset described in Sect. 7.2. *Left*: T versus β for BeeP, red contours ([68, 95, 99]%), and GCC, blue contours. *Middle and right*: GCC TEMP_CLUMP and BETA_CLUMP versus BeeP T and β, with 1 σ error bars. The black dashed line shows equality.

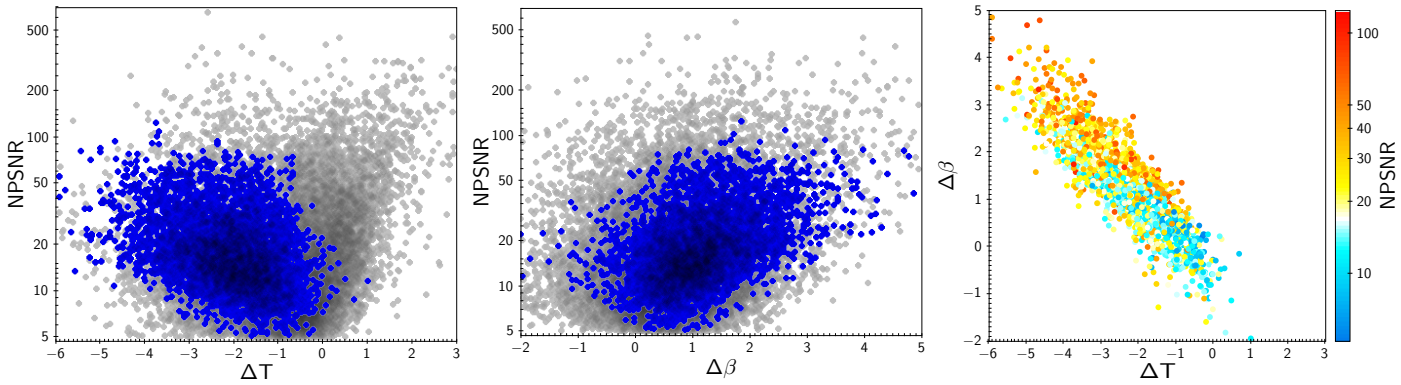

Fig. 27. *Left and middle*: detection significance level NPSNR versus normalized contrast ΔT (*left*) and $\Delta\beta$ (*middle*), defined in Eq. (19), for the set of sources in common between the BeeP and GCC catalogues (blue points), and for the entire BeeP/base catalogue (grey points). Correlations are seen for the common subset (shown in blue), but not for the entire BeeP/base catalogue (in grey). *Right*: ΔT versus ($\Delta\beta$) for the GCC common sample, with colour showing NPSNR. We see that ΔT and $\Delta\beta$ are highly correlated, and each is also correlated with NPSNR.

sources are of good quality according to GCC estimation (FLUX_QUALITY = 1), then the common set reduces to 5165 sources, with only 36 in the PCCS2. Of these, 73% (3757 sources) are in BeeP/base, and this is the set that we use for comparison.

Figure 26 shows a comparison between the BeeP MBB parameter estimates and their equivalent in GCC. There is good consistency between the two. Both methods show large uncertainty in T and β, which is not surprising, since we only have four frequencies and we are fitting a three-parameter model. There is a small positive bias in the GCC temperatures with respect to BeeP (+2.8% median, +3.2% mean). A small negative bias is also seen in the GCC spectral indices with respect to BeeP (−2.2% median and mean), which is also expected, considering the negative correlation between T and β.

In order to further assess consistency, we examine the difference between the two estimates, normalized by the combined uncertainty:

$$\delta\gamma \equiv \frac{\gamma_{\rm BeeP} - \gamma_{\rm GCC}}{\sqrt{\sigma^2_{\gamma_{\rm BeeP}} + \sigma^2_{\gamma_{\rm GCC}}}}\,, \tag{18}$$

where γ stands for either T or β. The dispersion of this quantity should be of order unity. Instead, we find that $\sigma_{\delta T} \approx 0.59$ and $\sigma_{\delta\beta} \approx 0.43$. However, our simulations already indicated that BeeP overestimates the error bars for both temperature and spectral index (see Table B.1). The extra deficit is probably the result of a positive correlation between the estimates of both methods, which arises from the fact that both use the same data.

We now turn to an assessment of the influence of the source-to-background temperature contrast on the estimation of significance, which is of interest because it is this contrast that drives the selection function of the GCC algorithm CoCoCoDeT (Planck Collaboration XXVIII 2016). We define

$$\Delta\xi \equiv \frac{\xi^{\rm source} - \xi^{\rm back}}{\sqrt{\sigma^2_{\rm source} + \sigma^2_{\rm back}}} \tag{19}$$

as the normalized contrast, where $\sigma_{\rm source}$ and $\sigma_{\rm backg}$ are the source and background 1 σ errors (assumed to be uncorrelated), and ξ can be either T or β.

We note that a contrast in the thermal properties of source and background translates into varying brightness ratios in each frequency channel. As a consequence, we expect that the MBB parameter source versus background contrast correlates with NPSNR via the BeeP likelihood. Indeed, when we limit the comparison to the GCC-matched subsample of BeeP/base shown in blue in the middle and left panels of Fig. 27, a positive

correlation appears between the contrast significance ΔT and $\Delta\beta$ (Eq. (19)) and the BeeP NPSNR. In addition, the right panel shows that the significance of the parameter recovery is positively correlated with the magnitude of the contrast. However, as may be seen from the grey points in the left and middle panels, when the entire BeeP/base catalogue is included, the contrast as estimated by BeeP shows only a mild correlation with NPSNR.

BeeP therefore recovers the source versus background contrast that the CoCoCoDeT algorithm uses to select GCC sources. However, the BeeP selection function is not limited to cold compact objects immersed in warm backgrounds, and provides a larger population of cold objects, including many that are not found in GCC. In fact, the BeeP base catalogue contains 11 145 cold objects ($T < 16$ K) in the region defined by the Galactic mask, as compared to 5489 with FLUX_QUALITY = 1 in the GCC.

7.3. Herschel H-Atlas catalogue (350 μm)

We have compared BeeP flux estimates with those in one field (GAMA15) of the *Herschel*-ATLAS catalogue (Valiante et al. 2016; Bourne et al. 2016). Because of the large disparity between the sensitivity and angular resolution of *Planck*-HFI and *Herschel*-SPIRE, we collated the catalogues by first selecting all H-ATLAS sources within a radius of 5′ around each *Planck* location. Then, for each *Planck* source we selected the brightest H-ATLAS source, which is not always the closest one. The two sets of flux densities (compared in Fig. 28) are, statistically, remarkably consistent. It is worth noting that the *Herschel* GAMA15 field follows quite closely the BeeP assumptions, in that the background is homogeneous and slowly-varying, and the foregrounds are well separated. All BeeP sources have SRCSIG values above 8.0 and low values of RELTH, except for one pair that is very close and mutually induces non-Gaussianity. BeeP errors are plotted as found in the catalogue (see Sect. 6.2.4).

7.4. Background estimates

As described in Sect. 6.3, the goal of the BeeP background analysis is not to provide an alternative characterization of *Planck*'s submillimetre diffuse background thermal properties, but rather to understand the impact of the background-foreground thermal contrast on the BeeP selection function. Nevertheless, it is interesting to check the validity of the BeeP background parameters. For this purpose, we have used the dust temperature and spectral-index maps from Planck Collaboration Int. XLVIII (2016), which have been extracted using the GNILC algorithm. We applied to these maps the procedure described in Sect. 6.3 used to compute the pairs $\{I_T, \sigma_{I_T}\}$ and $\{I_\beta, \sigma_{I_\beta}\}$, for direct comparison to the equivalent BeeP estimates.

Figure 29 shows a comparison of the MBB background parameters estimated by BeeP with those based on the GNILC dust component. The top row shows sources inside the PCCS2E Galactic mask (where dust emission is dominant). For these sources there is reasonably good agreement between both estimates. The bottom row shows the GNILC and BeeP background estimates in the $T-\beta$ plane for the same set of sources. The parameters (BeeP in blue and GNILC in red) are in good agreement; however, as the Galactic latitude increases, we start to see some disagreement (light blue points in the top row). Indeed, in the low-background PCCS2 region (middle row), the BeeP and GNILC estimates agree less well. In particular, the higher BeeP temperatures are significantly higher than the GNILC estimates. From Sect. 2.3 of Planck Collaboration XI (2014) we know that

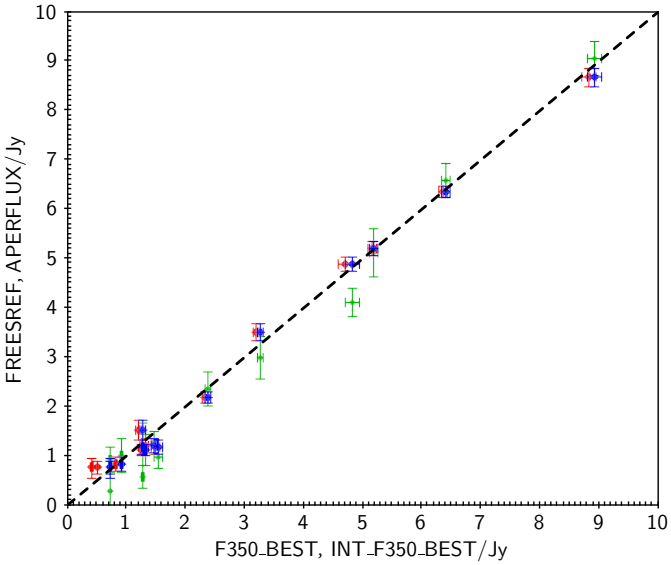

Fig. 28. Flux densities (FREESREF) of BeeP sources compared to those (F350_BEST) sources in the *Herschel* H-ATLAS GAMA15 field catalogue. The brightest H-ATLAS sources within 5′ of the corresponding *Planck* source are shown in red. The beam-weighted sums of all sources within the corresponding *Planck* 857 GHz beam are shown in blue. Green symbols give a similar comparison between the *Herschel* beam-weighted flux-density sum and the PCCS2 APERFLUX flux density. The BeeP errors are not corrected here (see Sect. 6.2.4).

in order to correctly fit the dust emission, we should first remove any other emission (CMB or CIB) and set map zero levels correctly. However, doing this would have biased the analysis of compact sources (e.g., the CMB is not determined at the location of the PCCS2+2E sources), and therefore the maps used as input to BeeP were not adjusted. This is the cause for the discrepancy with GNILC that we observe at high Galactic latitudes, where the relative weight of the CMB component, or even residual CIB anisotropies, is much higher. For this reason, the BeeP MBB parameter estimates of the background at high Galactic latitudes should not be taken to be good measures of the physical properties of dust emission. Their main purpose is to complement the characterization of the compact objects by adding a physical description of their embedding surroundings. Nevertheless, in regions of strong dust emission, the BeeP MBB parameter estimates are fairly good representations of T and β of dust in those regions.

8. Summary and conclusions

BeeP is a Bayesian algorithm that uses an assumed SED profile to combine observations of a source and its background at multiple frequencies, with the objective of evaluating the reliability of the source detection and estimating its physical properties. To implement BeeP, we developed a fast likelihood code, based on a simplified version of the data model, which overcomes the difficulties posed by the high data volume.

By applying the BeeP algorithm to the *Planck* 2015 maps at 353, 545, and 857 GHz, and the IRIS map at 3000 GHz, and assuming a dusty (MBB) SED, we constructed an extension to the *Planck* 857 GHz PCCS2 and PCCS2E single-channel catalogues, which provides new information on the reliability and physical properties of the sources in the catalogues. Since multiple frequencies are used, improved detection strength (NPSNR) is achieved. Our data model permits the construction of a

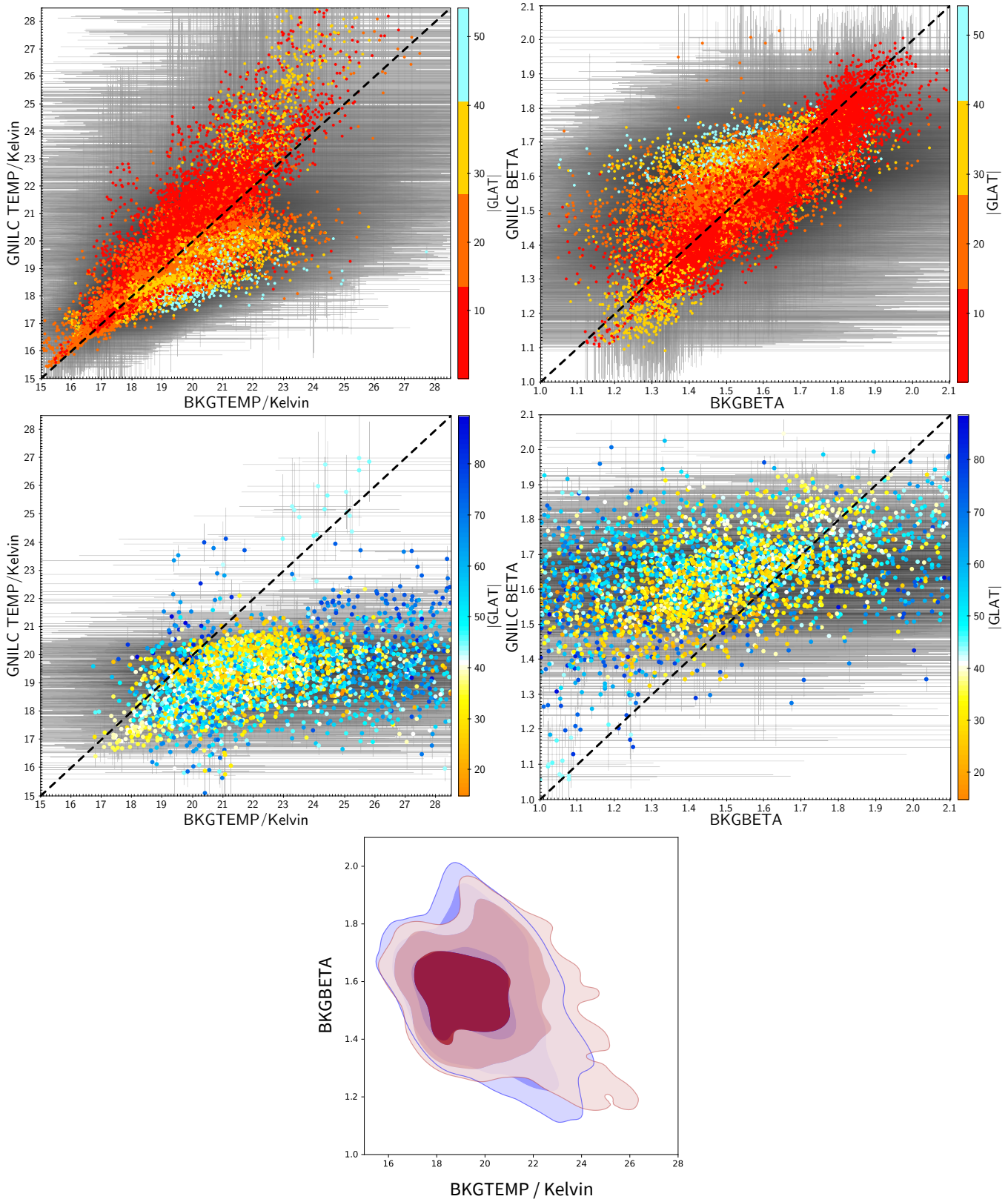

Fig. 29. Comparison of background T (*left*) and β (*right*) estimates from BeeP and GNILC. The colour of the points indicates the object's Galactic latitude, with grey lines being 1 σ error bars and the dashed black lines showing equality. *Top row*: sources inside the PCCS2+2E Galactic mask, i.e., regions with strong dust emission. *Middle row*: sources in the PCCS2 set, i.e., high-Galactic-latitude, dust-poor regions. *Bottom row*: same sources as the top row (i.e., inside the PCCS2+2E Galactic mask), but this time their T–β distribution. BeeP distribution contours are shown in blue ([68, 95, 99]%) and GNILC in red.

statistic to measure the local non-Gaussianity of the background (RELTH), which was used to correct NPSNR. The new significance statistic SRCSIG resulting from this process helps to separate foreground compact objects from the background, even in regions of strong and complex backgrounds such as cirrus and filaments. BeeP also determines a new effective source position that incorporates information at all frequencies used, which is not the same as what is provided in PCCS2+2E.

In addition to its determination of source reliability, BeeP provides a characterization of the thermal properties of each source and its background. As part of the BeeP output, we provide a figure for each source that displays the SED curves associated with the posterior parameter samples. We also provide the joint posterior distributions of the source MBB model parameters in "triangle" plots, which give, in each non-diagonal position, the marginalized bi-dimensional posterior PDF of the parameter samples defining the row and the column. The diagonal locations contain posterior marginalized distributions. The visualization of the posterior distributions enable a more complete understanding of the uncertainties associated with the source-parameter estimates. For instance, as expected, there is a strong correlation between the source MBB spectral index and its temperature.

For the sake of completeness, and to allow for a better comparison with PCCS2+2E data, we also extracted source parameters using a data model where the flux density of the source in each channel is a free independent parameter, the "Free" model. We also provide the MBB characterization of the background surrounding each source.

We tested the limits of our simplified data model and likelihood implementation, using an extensive battery of simulations, ranging from the fiducial, which closely follows the assumptions of the simplified data model, to the most realistic *Planck* simulations associated with the *Planck* 2015 data set, namely the full focal plane simulations (FFP8). To enhance confidence in the results, we also resorted to injection simulations, where a representative set of compact objects was injected into the real maps, extracted, and then cross-matched with the input catalogue. Analysis of the simulations allowed us to identify some data subsets where the optimality of the algorithm could not be guaranteed. We established a criterion of quality (EST_QUALITY), which can be used to filter out these anomalous sources.

The simulations were used to evaluate the effect of beam variations and ellipticity (which are not considered in our basic model) on parameter estimation. In particular, we find an 11% bias in the estimated flux densities, which we corrected in the output catalogue. Simulations also allowed us to determine that the uncertainties estimated by BeeP for flux densities and positions are unrealistically small for sources with very high values of NPSNR. We suggest procedures to correct these uncertainties for the small fraction of sources affected, but we did not apply them to the output catalogue – they should only be used if the rigorous statistical characterization of samples including those sources is required.

Based on our analysis, we define a reliable and accurate subset of PCCS2+2E (BeeP/base) containing 26 083 sources (54.1% of PCCS2+2E), of which 21 997 are in PCCS2E (50.8%). The estimated contamination level of this subset is between 5% and 10%. This, on its own, significantly improves the original PCCS2E, which contains no validated indicator of source reliability. Further imposing a criterion of contamination below 1%, BeeP still ranks 5077 compact objects in the PCCS2E as "good". Although the BeeP/base catalogue should be adequate for most purposes, we provide the relevant information needed by a user of our augmented version of PCCS2+2E to select a different subsample fitting specific scientific requirements (suggestions for selection criteria can be found in Sect. 5.6).

BeeP's selection function overlaps with that of the CoCoCoDeT extraction method used to generate the *Planck* Catalogue of Galactic Cold Clumps (GCC). The number of common objects between BeeP/base and the best quality detections in GCC (FLUX_QUALITY = 1) contains 3757 sources. We find good consistency in the thermal source parameters recovered by the two methods, considering the uncertainties in the estimation. The BeeP selection function is broader than that of GCC, even for the same range of temperatures, since the BeeP likelihood is not limited by the temperature contrast between a cold source and a warm background. The BeeP catalogue is, therefore, complementary to the GCC. For the GCC-selected sample, the BeeP parameter NPSNR (strength of the detection) is well-correlated with the source-to-background contrast.

The BeeP reference flux-density estimates (at 857 GHz) were also cross-checked against the PCCS2+2E estimates at 857 GHz and the *Herschel* GAMMA15-field catalogue at 350 μm. The match with the *Herschel* estimates is reasonably good when we include all sources within the *Planck* beam. The consistency with the PCCS2+2E flux-density estimates is also good, with only small biases, of known origin, but with some dispersion that is almost entirely the result of the large uncertainty in the PCCS2+2E aperture-photometry estimates. The BeeP flux-density uncertainty is significantly smaller (by a factor of 2) than that of the PCCS2+2E aperture-photometry estimates.

We also compared the BeeP estimates of background parameters against those of the GNILC temperature and spectral index dust maps. In those regions where dust is the dominant component, for instance within the PCCS2+2E Galactic masks, the agreement with the MBB thermal parameters is good. However, at high Galactic latitudes, where dust is no longer dominant and the CIB is strong, and, especially for the 353 GHz channel, the correlation is not as good, and the BeeP parameter estimates are less reliable.

In conclusion, we provide a new data set that characterizes the reliability and thermal properties of all sources in the PCCS2+2E. We expect this to greatly improve the utility of these catalogues. The results of this analysis will be made publicly available via the *Planck* Legacy Archive.

Acknowledgements. The Planck Collaboration acknowledges the support of: ESA; CNES and CNRS/INSU-IN2P3-INP (France); ASI, CNR, and INAF (Italy); NASA and DoE (USA); STFC and UKSA (UK); CSIC, MINECO, JA, and RES (Spain); Tekes, AoF and CSC (Finland); DLR and MPG (Germany); CSA (Canada); DTU Space (Denmark); SER/SSO (Switzerland); RCN (Norway); SFI (Ireland); FCT/MCTES (Portugal); and ERC and PRACE (EU). A description of the Planck Collaboration and a list of its members, indicating which technical or scientific activities they have been involved in, can be found at http://www.cosmos.esa.int/web/planck/planck-collaboration. We acknowledge support from the ESTEC Faculty Research Project Programme.

References

Barreiro, R. B., Sanz, J. L., Herranz, D., & Martínez-González, E. 2003, MNRAS, 342, 119
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Bouchaud, J.-P., & Potters, M. 2004, Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management, 2nd edn. (Cambridge: Cambridge University Press)
Bourne, N., Dunne, L., Maddox, S. J., et al. 2016, MNRAS, 462, 1714
Box, G., & Tiao, G. C. 1992, Bayesian Inference in Statistical Analisys (New York: John Wiley & Sons, Wiley-Interscience)
Cañameras, R., Nesvadba, N. P. H., Guery, D., et al. 2015, A&A, 581, A105
Carvalho, P., Rocha, G., Hobson, M. P., & Lasenby, A. 2012, MNRAS, 427, 1384

Casaponsa, B., Barreiro, R. B., Martínez-González, E., et al. 2013, MNRAS, 434, 796
Chen, X., Chary, R., Pearson, T. J., et al. 2016, MNRAS, 458, 3619
Clements, D. L., Braglia, F. G., Hyde, A. K., et al. 2014, MNRAS, 439, 1193
Combes, F., Rex, M., Rawle, T. D., et al. 2012, A&A, 538, L4
Csengeri, T., Urquhart, J. S., Schuller, F., et al. 2014, A&A, 565, A75
Delabrouille, J., Betoule, M., Melin, J.-B., et al. 2013, A&A, 553, A96
Egami, E., Rex, M., Rawle, T. D., et al. 2010, A&A, 518, L12
Eriksen, H. K., Jewell, J. B., Dickinson, C., et al. 2008, ApJ, 676, 10
Feroz, F., & Hobson, M. P. 2008, MNRAS, 384, 449
Foreman-Mackey, D. 2016, JOSS, 1, 24
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Goodman, J., & Weare, J. 2010, Commun. Appl. Math. Comput. Sci., 5, 65
Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759
Hobson, M. P., Rocha, G., & Savage, R. S. 2009, in Bayesian Methods in Cosmology, eds. M. P. Hobson, A. H. Jaffe, A. R. Liddle, P. Mukherjee, & D. Parkinson (Cambridge: Cambridge University Press), 167
Jaynes, E. T. 2004, Probability Theory: The Logic of Science (Cambridge: Cambridge University Press)
Low, F. J., Beintema, D. A., Gautier, T. N., et al. 1984, ApJ, 278, L19
Men'shchikov, A. 2013, A&A, 560, A63
Mitra, S., Rocha, G., Górski, K. M., et al. 2011, ApJS, 193, 5
Miville-Deschênes, M.-A., & Lagache, G. 2005, ApJS, 157, 302
Molinari, S., Schisano, E., Faustini, F., et al. 2011, A&A, 530, A133
Montier, L. A., Pelkonen, V.-M., Juvela, M., Ristorcelli, I., & Marshall, D. J. 2010, A&A, 522, A83
Nettke, W., Scott, D., Gibb, A. G., et al. 2017, MNRAS, 468, 250
Planck Collaboration ES 2018, The Legacy Explanatory Supplement https://www.cosmos.esa.int/web/planck/pla (ESA)
Planck Collaboration VII. 2011, A&A, 536, A7
Planck Collaboration XIX. 2011, A&A, 536, A19
Planck Collaboration IX. 2014, A&A, 571, A9
Planck Collaboration XI. 2014, A&A, 571, A11
Planck Collaboration XXVIII. 2014, A&A, 571, A28
Planck Collaboration I. 2016, A&A, 594, A1
Planck Collaboration VII. 2016, A&A, 594, A7
Planck Collaboration VIII. 2016, A&A, 594, A8
Planck Collaboration XII. 2016, A&A, 594, A12
Planck Collaboration XXVI. 2016, A&A, 594, A26
Planck Collaboration XXVII. 2016, A&A, 594, A27
Planck Collaboration XXVIII. 2016, A&A, 594, A28
Planck Collaboration I. 2020, A&A, 641, A1
Planck Collaboration Int. XXXIX. 2016, A&A, 596, A100
Planck Collaboration Int. XLVIII. 2016, A&A, 596, A109
Planck Collaboration Int. LIV. 2018, A&A, 619, A94
Riley, K., Hobson, M., & Bence, S. 2006, Mathematical Methods for Physics and Engineering: A Comprehensive Guide, 3rd edn. (Cambridge: Cambridge University Press)
Robert, C., & Casella, G. 2010, Introducing Monte Carlo Methods with R, Use R! (New York: Springer)
Schäfer, B. M., Pfrommer, C., Hell, R. M., & Bartelmann, M. 2006, MNRAS, 370, 1713
Stetson, P. B. 1987, PASP, 99, 191
Tegmark, M., & de Oliveira-Costa, A. 1998, ApJ, 500, L83
Trees, H. L. V. 2001, Detection, Estimation, and Modulation Theory. Part I (New York: John Wiley & Sons, Inc)
Valiante, E., Smith, M. W. L., Eales, S., et al. 2016, MNRAS, 462, 3146
Williams, J. P., Blitz, L., & Stark, A. A. 1995, ApJ, 451, 252

[1] AIM, CEA, CNRS, Université Paris-Saclay, Université Paris-Diderot, Sorbonne Paris Cité, 91191 Gif-sur-Yvette, France
[2] APC, AstroParticule et Cosmologie, Université Paris Diderot, CNRS/IN2P3, CEA/lrfu, Observatoire de Paris, Sorbonne Paris Cité, 10 Rue Alice Domon et Léonie Duquet, 75205 Paris Cedex 13, France
[3] African Institute for Mathematical Sciences, 6-8 Melrose Road, Muizenberg, Cape Town, South Africa
[4] Astrophysics Group, Cavendish Laboratory, University of Cambridge, J J Thomson Avenue, Cambridge CB3 0HE, UK
[5] Astrophysics & Cosmology Research Unit, School of Mathematics, Statistics & Computer Science, University of KwaZulu-Natal, Westville Campus, Private Bag X54001, Durban 4000, South Africa
[6] CITA, University of Toronto, 60 St. George St., Toronto, ON M5S 3H8, Canada
[7] CMDLABS, Cambridge Machines Deep Learning and Bayesian Systems Ltd, 22 Wycombe End, Beaconsfield, Buckinghamshire HP9 1NB, UK
[8] CNRS, IRAP, 9 Av. colonel Roche, BP 44346, 31028 Toulouse Cedex 4, France
[9] California Institute of Technology, Pasadena, CA, USA
[10] Computational Cosmology Center, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[11] Département de Physique Théorique, Université de Genève, 24, Quai E. Ansermet, 1211 Genève 4, Switzerland
[12] Département de Physique, École normale supérieure, PSL Research University, CNRS, 24 Rue Lhomond, 75005 Paris, France
[13] Departamento de Astrofísica, Universidad de La Laguna (ULL), 38206 La Laguna, Tenerife, Spain
[14] Departamento de Física, Universidad de Oviedo, C/ Federico García Lorca, 18, Oviedo, Spain
[15] Department of Astrophysics/IMAPP, Radboud University, PO Box 9010, 6500 GL Nijmegen, The Netherlands
[16] Department of Physics & Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, BC, Canada
[17] Department of Physics & Astronomy, University of the Western Cape, Cape Town 7535, South Africa
[18] Department of Physics, Gustaf Hällströmin katu 2a, University of Helsinki, Helsinki, Finland
[19] Department of Physics, Princeton University, Princeton, NJ, USA
[20] Dipartimento di Fisica e Astronomia G. Galilei, Università degli Studi di Padova, Via Marzolo 8, 35131 Padova, Italy
[21] Dipartimento di Fisica e Scienze della Terra, Università di Ferrara, Via Saragat 1, 44122 Ferrara, Italy
[22] Dipartimento di Fisica, Università La Sapienza, P. le A. Moro 2, Roma, Italy
[23] Dipartimento di Fisica, Università degli Studi di Milano, Via Celoria, 16, Milano, Italy
[24] Dipartimento di Fisica, Università degli Studi di Trieste, Via A. Valerio 2, Trieste, Italy
[25] Dipartimento di Fisica, Università di Roma Tor Vergata, Via della Ricerca Scientifica, 1, Roma, Italy
[26] European Space Agency, ESAC, Planck Science Office, Camino bajo del Castillo, s/n, Urbanización Villafranca del Castillo, Villanueva de la Cañada, Madrid, Spain
[27] European Space Agency, ESTEC, Keplerlaan 1, 2201 AZ Noordwijk, The Netherlands
[28] Gran Sasso Science Institute, INFN, Viale F. Crispi 7, 67100 L'Aquila, Italy
[29] HEP Division, Argonne National Laboratory, Lemont, IL 60439, USA
[30] Haverford College Astronomy Department, 370 Lancaster Avenue, Haverford, PA, USA
[31] Helsinki Institute of Physics, Gustaf Hällströmin katu 2, University of Helsinki, Helsinki, Finland
[32] INAF – OAS Bologna, Istituto Nazionale di Astrofisica – Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Area della Ricerca del CNR, Via Gobetti 101, 40129 Bologna, Italy
[33] INAF – Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, Padova, Italy
[34] INAF – Osservatorio Astronomico di Trieste, Via G.B. Tiepolo 11, Trieste, Italy
[35] INAF, Istituto di Radioastronomia, Via Piero Gobetti 101, 40129 Bologna, Italy
[36] INAF/IASF Milano, Via E. Bassini 15, Milano, Italy
[37] INFN – CNAF, Viale Berti Pichat 6/2, 40127 Bologna, Italy
[38] INFN, Sezione di Bologna, Viale Berti Pichat 6/2, 40127 Bologna, Italy
[39] INFN, Sezione di Ferrara, Via Saragat 1, 44122 Ferrara, Italy
[40] INFN, Sezione di Roma 2, Università di Roma Tor Vergata, Via della Ricerca Scientifica, 1, Roma, Italy
[41] Institut d'Astrophysique Spatiale, CNRS, Univ. Paris-Sud, Université Paris-Saclay, Bât. 121, 91405 Orsay Cedex, France
[42] Institut d'Astrophysique de Paris, CNRS (UMR7095), 98 bis Boulevard Arago, 75014 Paris, France

[43] Institute Lorentz, Leiden University, PO Box 9506, Leiden 2300 RA, The Netherlands
[44] Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
[45] Institute of Theoretical Astrophysics, University of Oslo, Blindern, Oslo, Norway
[46] Instituto de Astrofísica de Canarias, C/Vía Láctea s/n, La Laguna, Tenerife, Spain
[47] Instituto de Física de Cantabria (CSIC-Universidad de Cantabria), Avda. de los Castros s/n, Santander, Spain
[48] Istituto Nazionale di Fisica Nucleare, Sezione di Padova, Via Marzolo 8, 35131 Padova, Italy
[49] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA, USA
[50] Jodrell Bank Centre for Astrophysics, Alan Turing Building, School of Physics and Astronomy, The University of Manchester, Oxford Road, Manchester M13 9PL, UK
[51] Kavli Institute for Cosmology Cambridge, Madingley Road, Cambridge CB3 0HA, UK
[52] Laboratoire d'Océanographie Physique et Spatiale (LOPS), Univ. Brest, CNRS, Ifremer, IRD, Brest, France
[53] Laboratoire de Physique Subatomique et Cosmologie, Université Grenoble-Alpes, CNRS/IN2P3, 53 Rue des Martyrs, 38026 Grenoble Cedex, France
[54] Laboratoire de Physique Théorique, Université Paris-Sud 11 & CNRS, Bâtiment 210, 91405 Orsay, France
[55] Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[56] Low Temperature Laboratory, Department of Applied Physics, Aalto University, Espoo, 00076 Aalto, Finland
[57] Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85741 Garching, Germany
[58] Mullard Space Science Laboratory, University College London, Surrey RH5 6NT, UK
[59] NAOC-UKZN Computational Astrophysics Centre (NUCAC), University of KwaZulu-Natal, Durban 4000, South Africa
[60] National Centre for Nuclear Research, ul. L. Pasteura 7, 02-093 Warsaw, Poland
[61] Purple Mountain Observatory, No. 8 Yuan Hua Road, 210034 Nanjing, PR China
[62] SISSA, Astrophysics Sector, Via Bonomea 265, 34136 Trieste, Italy
[63] San Diego Supercomputer Center, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA
[64] School of Chemistry and Physics, University of KwaZulu-Natal, Westville Campus, Private Bag X54001, Durban 4000, South Africa
[65] School of Physics and Astronomy, Cardiff University, Queens Buildings, The Parade, Cardiff CF24 3AA, UK
[66] School of Physics and Astronomy, Sun Yat-sen University, 2 Daxue Rd, Tangjia, Zhuhai, PR China
[67] School of Physics, Indian Institute of Science Education and Research Thiruvananthapuram, Maruthamala PO, Vithura, Thiruvananthapuram 695551, Kerala, India
[68] Simon Fraser University, Department of Physics, 8888 University Drive, Burnaby BC, Canada
[69] Sorbonne Université, CNRS, UMR 7095, Institut d'Astrophysique de Paris, 98 bis bd Arago, 75014 Paris, France
[70] Sorbonne Université, Observatoire de Paris, Université PSL, École normale supérieure, CNRS, LERMA, 75005 Paris, France
[71] Space Science Data Center – Agenzia Spaziale Italiana, Via del Politecnico snc, 00133 Roma, Italy
[72] Space Sciences Laboratory, University of California, Berkeley, CA, USA
[73] The Oskar Klein Centre for Cosmoparticle Physics, Department of Physics, Stockholm University, AlbaNova 106 91, Stockholm, Sweden
[74] Université de Toulouse, UPS-OMP, IRAP, 31028 Toulouse Cedex 4, France
[75] Warsaw University Observatory, Aleje Ujazdowskie 4, 00-478 Warszawa, Poland

Appendix A: BeeP

In this appendix we provide a more detailed description of our algorithm, which we have named Bayesian Estimation and Extraction Package, and which we refer to as BeeP.

A.1. Characterization method

The Bayesian system of inference is an extension of deductive logic ({0 = false, 1= true} to a broader class of "degrees-of-belief" that consistently maps them into the real interval [0, 1] (Jaynes 2004, Chaps. 1 and 2). It associates those degrees-of-belief with conditional probabilities. So, if we represent the quantities we are interest in, like source position, flux, SED etc., by parameter vector $\boldsymbol{\Theta}$, the relevant question we can ask is: what is the joint probability distribution of our parameter vector $\boldsymbol{\Theta}$, given our data $\boldsymbol{d}$ and model assumptions H:

$$\Pr(\boldsymbol{\Theta}|\boldsymbol{d}, H). \tag{A.1}$$

It is possible to relate the quantity we are interested in with others that can be computed with the help of Bayes theorem:

$$\Pr(\boldsymbol{\Theta}|\boldsymbol{d}, H) = \frac{\Pr(\boldsymbol{d}|\,\boldsymbol{\Theta}, H)\,\Pr(\boldsymbol{\Theta}|H)}{\Pr(\boldsymbol{d}|H)}, \tag{A.2}$$

where $\Pr(\boldsymbol{\Theta}|\boldsymbol{d}, H)$ is the "posterior probability" distribution of $\boldsymbol{\Theta}$, $\Pr(\boldsymbol{d}|\,\boldsymbol{\Theta}, H) \equiv \mathcal{L}(\boldsymbol{\Theta})$ is the likelihood, $\Pr(\boldsymbol{\Theta}|H) \equiv \pi(\boldsymbol{\Theta})$ is the probability distribution of the variables of interest before considering the data, or the "prior" and $\Pr(\boldsymbol{d}|H)$ is the Bayesian "evidence", which, in this case, does not depend on any variable. Therefore, the evidence will only act as a normalizing constant and will be ignored. So, our main inference equation will read,

$$\Pr(\boldsymbol{\Theta}|\boldsymbol{d}, H) \propto \mathcal{L}(\boldsymbol{\Theta})\,\pi(\boldsymbol{\Theta}). \tag{A.3}$$

Once we have defined the likelihood and the prior functions, the parameter manifold $\boldsymbol{\Theta}$ is sampled using a Markov-chain Monte Carlo (MCMC) algorithm (Robert & Casella 2010). Choosing the right MCMC algorithm is still very much a matter of trial and error, since an optimal choice depends very much on the parameters manifold topology. In our case we expect a very heterogeneous manifold. Variables like temperature (T) and spectral index (β) are highly correlated and generate deep curved likelihood valleys, particularly for high signal-to-noise (S/N) ratio sources. The source flux density (S) and extension/radius (r) variables are expected to be correlated as well; however, the correlation is mostly linear and not very narrow. Additionally, the position vector variables (X, Y) are completely uncorrelated with all others. After reviewing several candidate algorithms, we chose `MCMC Hammer`[22]. which is currently popular in astrophysics and well adapted to sample from a likelihood manifold like ours. However, we did not use the available `python` code version because it did not show the required performance. A completely new sampler code was written in C++, based on the same algorithm (Goodman & Weare 2010). When running this code we chose to set all prior distributions to be uniform within a defined range[23].

[22] Foreman-Mackey et al. (2013)

[23] Given that we are only estimating the parameter posterior distributions, there is no need to define a precise and well-motivated range for the uniform distributions. We just need to make sure they are wide enough to not truncate a significant fraction of the likelihood volume for the physically possible parameter ranges (see Appendix A.1.3).

A.1.1. Likelihood

In Sect. 3.3, we have described how we build a model for *each* of the sources independently, $\boldsymbol{s}_j(\boldsymbol{\Theta}_j)$, which we combine with a model for the background, $\boldsymbol{b}_j(x)$, and the noise, $\boldsymbol{n}_j(x)$, in the neighbourhood of the source. Following the same principles, but extending to the full data set, the data model would now read:

$$\boldsymbol{d}(\boldsymbol{x}) = \sum_{j=1}^{N_s} \boldsymbol{s}_j(\boldsymbol{x}; \boldsymbol{\Theta}_j) + \boldsymbol{b}(\boldsymbol{x}) + \boldsymbol{n}(\boldsymbol{x}). \tag{A.4}$$

where for convenience we concatenated the individual background and noise quantities into the full sky $\boldsymbol{b}(\boldsymbol{x})$ and $\boldsymbol{n}(\boldsymbol{x})$ quantities. Given the assumptions described in Sect. 3.1 the likelihood representing *all* compact objects in the map is

$$\mathcal{L}(\boldsymbol{\Theta}) = \frac{\exp\left\{-\frac{1}{2}\left[\boldsymbol{d} - \widehat{\boldsymbol{b}} - \boldsymbol{s}(\boldsymbol{\Theta})\right]^{\mathrm{t}} \mathsf{N}^{-1}\left[\boldsymbol{d} - \widehat{\boldsymbol{b}} - \boldsymbol{s}(\boldsymbol{\Theta})\right]\right\}}{(2\pi)^{N_{\rm pix}/2}\,|\mathsf{N}|^{1/2}}, \tag{A.5}$$

where $\boldsymbol{d}$ is the data (pixels), $\widehat{\boldsymbol{b}}$ is the generalized background ($\boldsymbol{b} + \boldsymbol{n}$) and N is the generalized background covariance matrix. For compactness, we merged all individual source parameters ($\boldsymbol{\Theta}_j$) into $\boldsymbol{\Theta}$. N is a huge matrix $N_{\rm pix} \times N_{\rm pix}$, where $N_{\rm pix} \sim 50\,000\,000$. Any brute force attempt to evaluate the likelihood will undoubtedly be frustrated by the sheer magnitude of the problem. This is where we take advantage of the homogeneity condition. Since N is the covariance matrix of a homogeneous Gaussian random field, by definition it is "circulant". Therefore, when represented in Fourier space it becomes diagonal. Performing this transformation, the full-sky source signal (Eq. (A.4)) in Fourier space reads,

$$\widetilde{\boldsymbol{s}}(\boldsymbol{\eta}; \boldsymbol{\Theta}) = \widetilde{\boldsymbol{B}}(\boldsymbol{\eta}) \sum_{j=1}^{N_s} A_j \boldsymbol{f}(\boldsymbol{\phi}_j)\,\widetilde{\tau}(-\boldsymbol{\eta}; \boldsymbol{a}_j) \mathrm{e}^{\mathrm{i}2\pi\boldsymbol{\eta}\cdot\boldsymbol{X}_j}, \tag{A.6}$$

where the vector $\widetilde{\boldsymbol{B}}(\boldsymbol{\eta})$ contains the Fourier transform of the beam at each frequency[24], $\boldsymbol{f}(\boldsymbol{\phi}_j)$ contains the Fourier transform of the emission coefficients at each frequency, and $\widetilde{\tau}(\boldsymbol{\eta}; \boldsymbol{a})$ is the Fourier transform of the template for an unconvolved object at the origin, characterized by the shape parameters $\boldsymbol{a}$.

We now consider the likelihood of the "no-source" model $\mathcal{L}_0$, i.e., when A, the source amplitude is equal to 0. $\mathcal{L}_0$ is a constant, since it does not contain any parameter. By taking the logarithm of the $\mathcal{L}(\boldsymbol{\Theta})/\mathcal{L}_0$ ratio, we reach a likelihood expression that reads

$$\begin{aligned}
&\ln\left[\mathcal{L}(\boldsymbol{\Theta}) L_0\right] \\
&= \sum_{j}^{N_s} \left\{ A_j \mathcal{F}^{-1}\left[\mathcal{P}_j(\boldsymbol{\eta})\widetilde{\tau}(-\boldsymbol{\eta}; \boldsymbol{a}_j)\right]_{\boldsymbol{X}_j} - \tfrac{1}{2} A_j^2 \sum_{\boldsymbol{\eta}} Q_{jj}(\boldsymbol{\eta}) |\widetilde{\tau}(\boldsymbol{\eta}; \boldsymbol{a}_j)|^2 \right\} \\
&\quad - \sum_{i>j}^{N_s} \left\{ A_i A_j \mathcal{F}^{-1}\left[Q_{ij}(\boldsymbol{\eta})\widetilde{\tau}(\boldsymbol{\eta}; \boldsymbol{a}_i)\widetilde{\tau}(-\boldsymbol{\eta}; \boldsymbol{a}_j)\right]_{\boldsymbol{X}_i - \boldsymbol{X}_j} \right\},
\end{aligned} \tag{A.7}$$

where $\mathcal{F}^{-1}[\ldots]_{\boldsymbol{x}}$ denotes the inverse Fourier transform of the quantity in brackets, evaluated at the point $\boldsymbol{x}$. We have also defined the following quantities:

- *Point source response* (or beam shape, i.e., how the data responds to the presence of a point source), $(\boldsymbol{\psi}_i)_\nu = \widetilde{B}_\nu(\boldsymbol{\eta})(\boldsymbol{f}_i)_\nu$, with ν labelling frequency channels;

[24] The beam transfer function is also convolved with the pixel window function at each frequency. In this particular case the pixel window function does not change across maps.

– *Information source* (on the point sources),
$\mathcal{P}_j(\boldsymbol{\eta}) \equiv \widetilde{\boldsymbol{d}}^{\mathsf{T}}(\boldsymbol{\eta})\mathsf{N}^{-1}(\boldsymbol{\eta})\boldsymbol{\psi}(\boldsymbol{\eta})$;
– *Point-source flux precision matrix*
$Q_{ij}(\boldsymbol{\eta}) \equiv \widetilde{\boldsymbol{\psi}}_i^{\mathsf{T}}(\boldsymbol{\eta})\mathsf{N}^{-1}(\boldsymbol{\eta})\boldsymbol{\psi}_j(\boldsymbol{\eta})$,

where N is the covariance matrix represented in Fourier space.

Let us take a closer look at the second term of Eq. (A.7), the "cross-term", where multiple sources interact. To help clarify the physical meaning of this expression let us write it for the "pure" point source, $\tau(\boldsymbol{x}, \boldsymbol{a}) = \delta(\boldsymbol{x})$:

$$\sum_{i>j} A_i A_j \mathcal{F}^{-1}[Q_{ij}(\boldsymbol{\eta})]_{X_i - X_j}. \tag{A.8}$$

For simple, uncorrelated backgrounds, $Q_{ij}(\boldsymbol{\eta})$ contains just linear combinations of the instrument beams at each frequency channel. The condition for this expression to become small, when compared with the rest of the likelihood, is the common assumption in astronomy that beam blending effects are negligible. If the sources are well separated such that $|\boldsymbol{X}_i - \boldsymbol{X}_j|$ is large enough for the multi-channel equivalent beam to have died out, then the contribution of Eq. (A.8) to the likelihood may be safely dropped[25]. So, assuming

$$\sum_{i>j}^{N_s} \left\{ A_i A_j \mathcal{F}^{-1}\left[Q_{ij}(\boldsymbol{\eta})\,\widetilde{\tau}(\boldsymbol{\eta}; \boldsymbol{a}_i)\,\widetilde{\tau}(-\boldsymbol{\eta}; \boldsymbol{a}_j)\right]_{X_i - X_j} \right\} \approx 0, \tag{A.9}$$

is nothing more than a generalization of the common assumption that objects are well separated so that we can ignore object blending effects caused by the beam[26]. We are left with the likelihood expression, which we sample from

$$\ln\left[\mathcal{L}(\boldsymbol{\Theta})/\mathcal{L}_0\right] = \sum_{j}^{N_s} \left\{ A_j \mathcal{F}^{-1}\left[\mathcal{P}_j(\boldsymbol{\eta})\widetilde{\tau}(-\boldsymbol{\eta}; \boldsymbol{a}_j)\right]_{X_j} - \tfrac{1}{2} A_j^2 \sum_{\boldsymbol{\eta}} Q_{jj}(\boldsymbol{\eta}) |\widetilde{\tau}(\boldsymbol{\eta}; \boldsymbol{a}_j)|^2 \right\}. \tag{A.10}$$

Equation (A.10) is extremely convenient from a computational point of view. The likelihood ratio, when neglecting the blending effects, becomes the sum of the individual contributions from each source. This allows one to use a very convenient "one source at a time" approach.

If $\widehat{\boldsymbol{\Theta}}_j$ is the set of parameter values that maximizes the likelihood ratio (Eq. (A.10)) for source j then

$$\ln\left(\frac{\mathcal{L}(\widehat{\boldsymbol{\Theta}}_j)}{\mathcal{L}_0}\right) = \tfrac{1}{2} \sum_{\boldsymbol{\eta}} \widehat{Q}_{jj}(\boldsymbol{\eta}) |\widetilde{\tau}(\boldsymbol{\eta}; \widehat{\boldsymbol{a}}_j)|^2 \widehat{A}_j^2 = \tfrac{1}{2} \widehat{\mathsf{NPSNR}}_j^{\,2}, \tag{A.11}$$

where we have defined the quantity $\mathcal{R}$ as the Neyman–Pearson S/N ratio, (corresponding to NPSNR in the catalogue). This variable is a function of the likelihood ratio, hence "Neyman–Pearson", but since

$$\sum_{\boldsymbol{\eta}} \widehat{Q}_{jj}(\boldsymbol{\eta}) |\widetilde{\tau}(\boldsymbol{\eta}; \widehat{\boldsymbol{a}}_j)|^2 = \frac{1}{\sigma^2}, \tag{A.12}$$

where σ^2 is the variance of the likelihood-ratio background field, it is also a signal-to-noise ratio, or the detection significance level (i.e., "how many sigma" this detection is). As we noted in Sect. 6.2.4, if all our assumptions hold and all source parameters were known except the amplitude (A), then NPSNR would indeed be the inverse of the fractional error on amplitude $A/\Delta A$.

A.1.2. Source-detection significance evaluation: dealing with the deviations from the data model

Although much of this has already been described in Sect. 3.3, we repeat a brief discussion of the evaluation of source significance here, in order to preserve the continuity of Appendix A.

There are two main data features that break the assumptions in our data model:
– background non-Gaussianity;
– localized structures.

It is well known that diffuse emission from dust, the main background component, is highly non-Gaussian. One may argue that because we are combining data from several channels, that increases the data volume and because of the central limit theorem the statistics should converge to Gaussian. Unfortunately, this is only true close to the mode of the distribution. But detection is all about the positive tail of the background distribution (see e.g., Fig. 3), and in this case the non-Gaussianity only decreases very slowly when more data are added (Bouchaud & Potters 2004, Chap. 2). However, an even larger problem comes from localized structures such as cirrus[27]. The likelihood ($\mathcal{L}(\widehat{\boldsymbol{\Theta}})$) of a cirrus cloud being confused for a source is small, given that cirrus is rather poorly described as a compact source. On the other hand, the likelihood of cirrus being a homogeneous Gaussian random field $\mathcal{L}_0$ is also very small, since by definition these structures do not behave as a homogeneous random field. So, by looking at Eq. (A.11), one can see that the source significance indicator

$$\mathsf{NPSNR} \propto \sqrt{\ln\left(\mathcal{L}(\widehat{\boldsymbol{\Theta}})/\mathcal{L}_0\right)},$$

might indeed create a strong positive tail event when a cirrus structure is present, even in the absence of a genuine source, and this might be taken (erroneously) to be an object of interest. It can be shown that, if all our assumptions hold, under the "null" hypothesis of our model (i.e., "only background is present") the following field is a white-noise unitary ($\sigma = 1$) Gaussian random field in pixel space ($\boldsymbol{X}$):

$$\frac{\mathcal{F}^{-1}\left[\widehat{\mathcal{P}}(\boldsymbol{\eta})\widetilde{\tau}(-\boldsymbol{\eta}; \widehat{\boldsymbol{a}})\right]_X}{\sqrt{\sum_{\boldsymbol{\eta}} \widehat{Q}(\boldsymbol{\eta}) |\widetilde{\tau}(\boldsymbol{\eta}; \widehat{\boldsymbol{a}})|^2}}. \tag{A.13}$$

Then if we added a point source to the centre of this perfect background we would introduce significant outliers in the positive tail of the distribution. Let us now assume that the positive outlier pixels created by the source are no more than a small fraction of the total number of pixels (α). Then using the quantile definition one would expect that

$$\int_{-\infty}^{\mathsf{RELTH}} \frac{\exp\left[-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right]}{\sqrt{2\pi}\sigma} \mathrm{d}x = 1 - \alpha, \tag{A.14}$$

where RELTH (reliability threshold) is the $1 - \alpha$ distribution quantile. RELTH can be read from the actual field histogram and

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

[26] Given *Planck*'s sensitivity, the surface density of sources is such that this condition holds well, except for in the Galactic plane.

[27] Cirrus is not the only type of localized feature. Extended sources that were identified as compact objects in the PCCS2, but where the actual positions were off the centre, also appear like localized artefacts.

then Eq. (A.14) solved for σ. If the remaining $1-\alpha$ pixels follow a unitary Gaussian distribution then $\sigma = 1$. However because of the enlarged distribution tails induced by the localized features and the background non-Gaussianity, σ is expected to be larger. Using simulations, we have verified that the number of outlier pixels created by the source is less than 5% of the total, so we use $\alpha = 5\%$.

Solving Eq. (A.14), σ is equal to

$$\sigma = k\ \mathsf{RELTH}, \tag{A.15}$$

where k is a pure numerical constant given by

$$k = \frac{1}{\sqrt{2}\,\mathrm{erfc}^{-1}(2\alpha)}, \tag{A.16}$$

where erfc^{-1} is the inverse complementary error function. We finally define the "source significance" estimator as

$$\mathsf{SRCSIG} = \frac{1}{k}\,\frac{\mathsf{NPSNR}}{\mathsf{RELTH}}, \tag{A.17}$$

where k is a pure numerical constant given by Eq. (A.16). This value is the same for all sources. If the histogram of the field given in Eq. (A.13) is Gaussian, then $\sqrt{2}\,\mathrm{erfc}^{-1}(2\alpha) = \mathsf{RELTH}$ by definition and $\mathsf{SRCSIG} = \mathsf{NPSNR}$. If our assumptions hold then, as predicted, NPSNR is the detection significance. However when there is non-Gaussianity in the background, either from diffuse components or localized features, then RELTH increases and a penalty is applied to the Gaussian criterion. This criterion becomes relaxed for high galactic latitudes away from cirrus where the homogeneity and Gaussian assumptions hold well, while in the neighbourhood of the Galactic plane or inside cirrus structures it becomes mores stringent to avoid false positives induced by the non-Gaussianity of the background[28].

A.1.3. Priors

We have tried to choose "non-informative" priors, constructing them such that the "maximum a posteriori" (MAP) estimator of any quantity depends exclusively on the current data set. One way of expressing this condition is that, when changing the data, the likelihood shape remains unchanged and only its location in the parameter space changes (Box & Tiao 1992, Chap. 1). Source position and amplitude are "location" parameters, at least within small ranges around the likelihood maxima. So all associated priors will be taken as uniform. The same cannot be said about the source extension parameter EXT (see definition in Sect. 6.2.2), which is a "hybrid" parameter that shifts and scales the likelihood (Carvalho et al. 2012). To improve the accuracy of the estimates, we use a "trick" (see Appendix A.2.3) that makes the objects always appear as if they were slightly extended. The prior on EXT should behave as $\pi(\mathsf{EXT}^{-2})$, and this function varies slowly for values of EXT away from 0. Since we target compact sources (i.e. close to beam-sized), we are able to select a narrow range (between 0.46 and 2.6 pixels), which allows us to replace the functional prior with a uniform one. This trick simplifies the problem without biasing the estimate of the value that maximizes the likelihood ($\widehat{\mathsf{EXT}}$).

Regarding the source brightness parameters (flux and spectral index), Eriksen et al. (2008) claim that using uniform priors instead of the Jeffreys non-informative priors creates a strong bias on the spectral index estimate. However, we have carried out an extensive battery of simulations to test this claim, and failed to find such a bias. Therefore, for simplicity we have kept the uniform prior distribution. It is important to keep the range of priors large enough to properly explore and characterize uncertainties. The ranges we selected bracket widely physically motivated values $\{\beta \in [0, 7]; T \in [3, 150]\,\mathrm{K}\}$[29]. We note, however, that the resulting range of values (see Fig. 16) is consistent with physically reasonable values and that the error bars do extend to much wider ranges.

A.1.4. Covariance matrix estimation: cross-correlation factor

The background cross-power spectrum matrix N is a critical part of the likelihood and our data model assumes we know its true value[30]. However, as it is not known a priori, an estimate must be computed. There are at least two completely different ways of tackling this problem. One way is by using theoretical models for each of the background components (for diffuse dust emission models see Planck Collaboration Int. XLVIII 2016 and Schäfer et al. 2006 for their application to the estimation of background cross-covariance). However powerful, this technique assumes full-sky statistical isotropy. A quick look at *Planck* maps immediately shows that these conditions are severely broken and hence the models are a sub-optimal approximation to real data.

A different approach (and the one we take) is to split the sky into small fields, where the isotropy conditions apply fairly well, and estimate the cross-power spectrum directly from the data. This method is not without its problems. For instance the data is one single realisation of the random process and not the ensemble average. In order to improve the estimation quality of the background covariance, we have developed a method based on the work of Bouchaud & Potters (2004, see Chap. 9) for time series. Expressing the problem in Fourier space allows us to treat each pixel *of the same channel*, or Fourier spatial mode, independently. However, each Fourier mode in channel k ($(\eta_i, \zeta_j)_k$) is correlated with the same Fourier mode in channel l ($(\eta_i, \zeta_j)_l$). So, assuming that each spatial Fourier mode (η_i, ζ_j) is one datum, and that we have N channels (N_{ch}), the covariance estimation quality factor for one single Fourier mode is given by

$$Q \equiv N_{\mathrm{rel}}/N_{\mathrm{ch}}, \tag{A.18}$$

where N_{rel} is the number of realisations of that particular Fourier mode. However, since for each patch we have one single realisation of the background $N_{\mathrm{rel}} = 1 \Rightarrow Q = 1/4$, an extremely low value. So, a simple estimate will be nothing but noise, as intuition would have told us. Assuming the process is ergodic and the field is homogeneous, it is possible to replace the ensemble average by a spatial average. So, we enlarge the patch to 16 times its initial area around the targeted source, (see the "field" definition in Appendix A.2.2 and Fig. A.3 for more details) and we average each background Fourier mode $(\eta_i, \zeta_j)_k$ over the same mode in different sub-regions (see "patch"). Since we now have 16 realisations of each individual Fourier mode, we have boosted Q to approximately 4[31]. This is already a reasonable estimation quality factor. However, the covariance matrix only enters the likelihood via its inverse. Even a small error in the estimation

[28] See Appendices A.2.3 and B.1 for the practicalities of applying Eq. (A.17).

[29] We allow β to go down to 0, to accommodate for flat spectra synchrotron sources.

[30] N is a set of 4×4 matrices, one for each pixel.

[31] It would have been exactly 4 if the "field" were perfectly homogeneous.

might render the inversion unstable. We have therefore decomposed the covariance matrix into two components:

$$\begin{bmatrix} \sigma^2_{353} & 0 & 0 & 0 \\ 0 & \sigma^2_{545} & 0 & 0 \\ 0 & 0 & \sigma^2_{857} & 0 \\ 0 & 0 & 0 & \sigma^2_{\rm IRIS} \end{bmatrix} + \phi \begin{bmatrix} 0 & kl_{353-545} & kl_{353-857} & kl_{353-\rm IRIS} \\ . & 0 & kl_{545-857} & kl_{545-\rm IRIS} \\ . & . & 0 & kl_{857-\rm IRIS} \\ . & . & . & 0 \end{bmatrix}, \tag{A.19}$$

where the matrix on the left is the "independent" component, the matrix on the right is the "systemic" component and ϕ is the "cross-correlation factor" ($\phi \in [0, 1]$). The independent component would be the covariance matrix if we neglected all cross-correlations between the same spatial Fourier mode in different channels. The systemic component is obtained as:

$$\sum_{i=1}^{N} \gamma_j \boldsymbol{V}_j \boldsymbol{V}_j^{\mathsf{T}}, \tag{A.20}$$

where γ_j and $\boldsymbol{V}_j$ are the eigenvalues and eigenvectors of the Karhunen–Loève decomposition of the Fourier mode (η_i, ζ_j) covariance matrix. Firstly, we start by sorting the KL eigenmodes in decreasing order of the respective eigenvalue. Then we include up to N ($\in [1, 4]$) eigenmodes in the sum of Eq. (A.20). N, the cut-off, is given by the theory of random covariance matrices (Bouchaud & Potters 2004, Chap. 9). After forming the systemic matrix using Eq. (A.20), we set the diagonal terms to 0.

Let us now inspect the two extreme cases of the cross-correlation factor:

- $\phi = 0$,
  the covariance becomes reduced to the independent only, which is equivalent to ignoring all cross-channel correlations;
- $\phi = 1$,
  we are including the cross-correlation between channels to its full extent only neglecting the modes that are severely contaminated by noise.

The behaviour of the background covariance-matrix estimator may be fine tuned using ϕ. A low value of ϕ improves the quality factor of the estimation at the cost of ignoring a portion of the signal, namely the inter-channel cross-correlation. A high value indicates inclusion of complete data information at the cost of lower estimation quality. As we shall see in Appendix B.2, the "cross-correlation factor" (ϕ) has proved to be important in the extraction of an unbiased $\{\beta, T\}$ set.

A.1.5. Incomplete modelling of the data: systematics

The determination of the covariance matrix of the cross-power spectra is an approximation, and any potential mis-estimation is not being propagated into the source parameter errors. However, we know that the dynamic range of source flux density and source-detection S/N is enormous, ranging from close to zero to the thousands. For most of the catalogue sources, the effect of the covariance-matrix estimation error is masked by the intrinsic uncertainty on the source parameters. However, for the most significant sources the covariance-matrix estimation error is likely to be the dominant effect, and for these we are missing a critical component of the uncertainty in the source parameters.

The rigorous and complete way of modelling the problem would be to include the uncertainty of the covariance-matrix coefficients as sampling variables in our problem and consider the joint likelihood. The inverse covariance-matrix coefficients[32] are distributed according to a Wishart distribution (Box & Tiao 1992, Chap. 8). In principle we could add this contribution to the source parameters and sample from the joint likelihood. However, we are dealing with seven or eight source parameters, and adding the inverse covariance likelihood would increase that number to more than 10 000. Sampling from tens of thousands of parameters would slow down the code to the point where it would no longer be possible to tackle a catalogue with more than 40 000 sources. Therefore we do not implement such a scheme. As a consequence, as we go up in source significance, some of the estimated parameter uncertainties will keep artificially decreasing, whereas in reality they should saturate at some minimum level. This effect particularly concerns the estimates of source location and flux density. In Appendix B we describe a wide range of simulations on which we have tested the limits of our approximation on several parameters, and suggest ways to correct this shortcoming.

A.2. Algorithm implementation

As described earlier, the maps that we use as inputs are *Planck* 2015 data at 353, 545, and 857 GHz, plus IRIS data at 3000 GHz. Here we describe some details of how we treat these data.

A.2.1. Masks and map sets

BeeP creates two types of masks, which are applied to the input maps to generate two types of map sets:

- IRIS
  The "IRIS" mask (see Fig. A.1) flags the regions on the IRIS map where there is incomplete data either because those regions were not observed or they contain compact sets of "ill-conditioned" pixels. The total area of this mask is about 3.3% of the full sky. There are 650 PCCS2+2E sources (1.4%) that are located within the IRIS mask. It is very difficult to constrain the emission temperatures using *Planck* data only; therefore objects positioned inside the IRIS mask are flagged in the catalogue as being of lower quality (see Table 1). The IRIS mask is applied to the input maps to provide a set of foreground maps (see Fig. A.2). All likelihood elements (except the background covariance) will be estimated using this data set. All injection, and non-injection, simulations only employ the foreground maps data set.

- Background
  The sole purpose of the "background" mask (see Fig. A.1) is to help in the removal of compact objects in order to create a set of "background" only maps. To construct this mask, firstly we merge all sources contained in the 353–857 GHz PCCS2+2E catalogues. We assume that this set of catalogues provides an almost complete sample at the sensitivities we are aiming for[33]. Then for every source in the merged catalogue we mask all the pixels inside a circle of 7′ radius. The 7′ radius was chosen to provide a good balance between an effective source brightness removal and, especially at low Galactic latitudes where the density of sources is very high,

[32] It is the inverse covariance that is part of our likelihood.

[33] The small excess of objects that are not part of these catalogues will then add to the background fluctuation levels and to its non-Gaussianity. This would only make our acceptance statistic even more conservative.

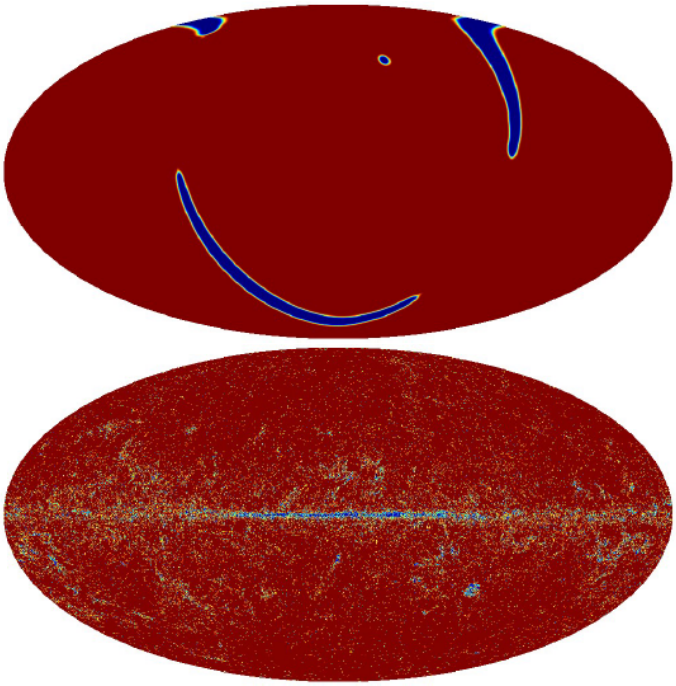

Fig. A.1. Masks used in our analysis. *Upper panel*: IRIS mask. The blue regions were not observed by IRAS or contain compact sets of "bad" pixels (3.3% of the sky). *Lower panel*: background mask. The masked regions were later inpainted by diffusing the hole boundary pixels into the interior.

to preserve the statistical properties of the background (see Fig. A.2). The total background area masked are is 4.3%. The background mask is applied to the input maps to provide a set of background maps (see Fig. A.2). The masked regions are then inpainted[34]. The main purpose of this set of maps is the evaluation of the background covariance matrix (cross-power spectrum).

A.2.2. Projection into flat fields

For each source in the PCCS2+2E, a flat area of size 3°.69×3°.69, (129 × 129) pixels, centred on the source position and obtained using a gnomonic projection, is cut from each *Planck* (353–857 GHz) and IRIS map. We repeat the procedure for the background map set and the IRIS mask (see Fig. A.3). We call these projected square maps "fields". Each individual pixel in each field is uniformly over-sampled by a factor of 25 to minimize resampling artefacts that could result from the overlap between the map and field grids[35]. Each sample is computed by bi-linearly interpolating the map pixels. The combination of the oversampling and the interpolation operations also smooths the map brightness. That effect is accounted for by adding a pixel-window correction to the effective beams.

The field is then divided into 49 (7×7) overlapping "patches" of 33 × 33 pixels, as shown in Fig. A.3[36]. The cross-power spectrum is computed in each of the patches and then averaged over all patches in a given field (see Appendix A.1.4). The IRIS mask is used to down-weight individual cross-spectrum Fourier modes according to the number of pixels removed by the mask. Since there is overlap between patches, the quality ratio improvement (Eq. (A.18)) is limited to about a fctor of 16. Finally, the likelihood/posterior is computed only using the central patch (Fig. A.3 in red), which is centred on the PCCS2+2E target object's original position (in yellow).

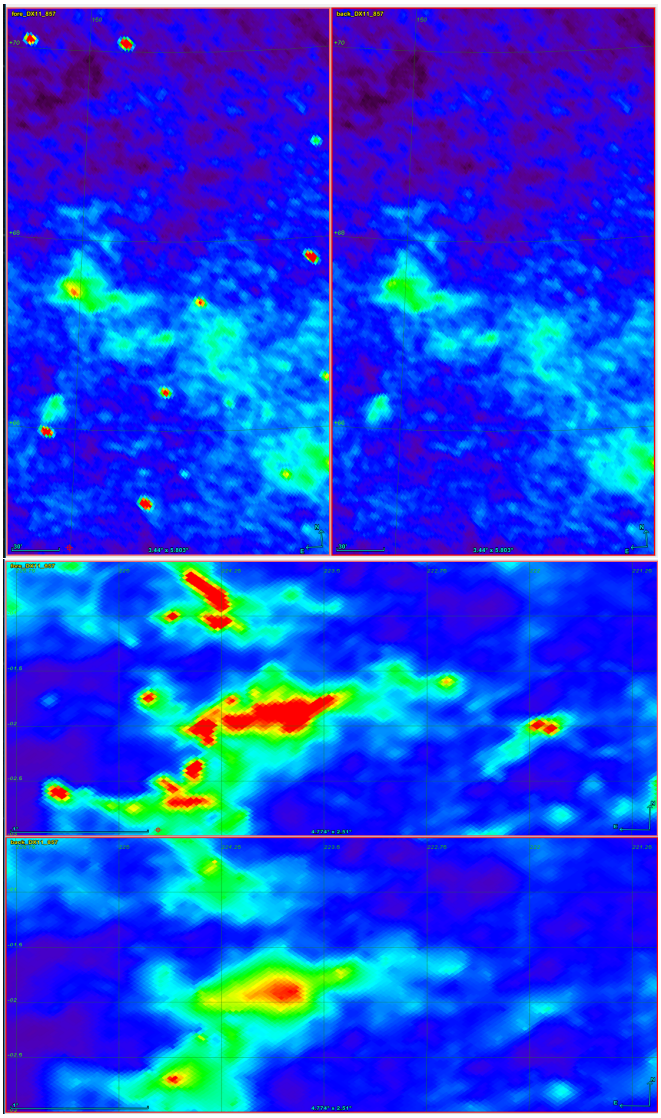

Fig. A.2. Masking and inpainting effects. Each of the *top panels* shows a small (3°.4×5°.8) high Galactic latitude patch cut from *Planck*'s 857 GHz map. The brightness-colour mapping is the same for both panels. The *left panel* is from the foreground map and the *right* from the background. For this low spatial density field, the "mask+inpaint" method recovers the background brightness map very accurately. The *two lower panels*, with foreground above and background below, show a very bright low Galactic latitude region (4°.8×2°.5). In this region of high spatial density of sources the "mask+inpaint" process is much less accurate and some degradation of the background can be seen.

A.2.3. Running the likelihood

For our likelihood runs we have set the "cross-correlation factor" (see Appendix A.1.4) ϕ to 10%. This value was selected using

[34] BeeP reports the percentage of pixels that were changed by the inpainting routine in a field labelled INPIX.

[35] This is equivalent to a field/patch pixel grid with a resolution 25 times greater than that of *Planck* HFI, which was later downgraded back to *Planck*'s original map resolution.

[36] This procedure is a simple 2-d extension of the quite common equivalent method in time series. If a time series is ergodic and stationary, the ensemble average may be replaced by a time/space average.

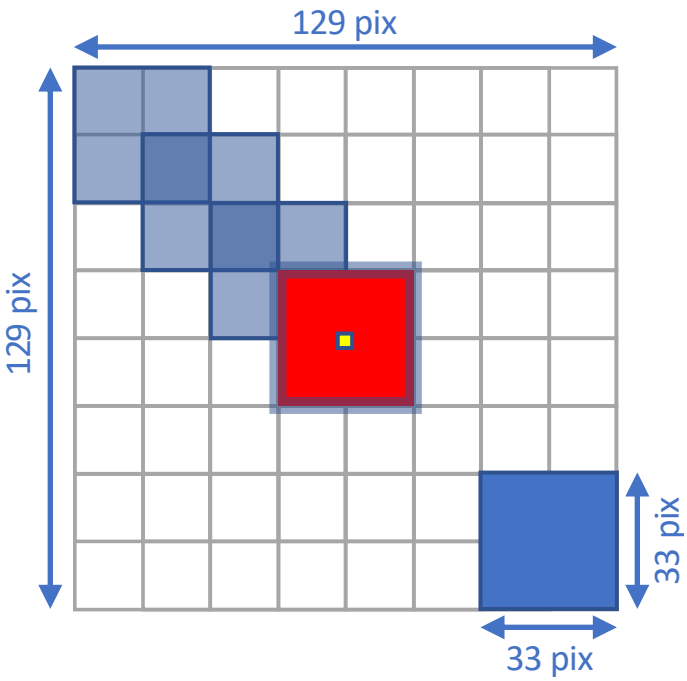

Fig. A.3. Schematic (not drawn to scale) showing parts of a flat "field" (129 × 129 pixels). The covariance matrix is computed at each of the large squares, or "patches" (33×33 pixels), from the "background" map, and then averaged over them. There are 49 overlapping patches (7×7) in each field. These are laid out as shown in the figure. The full likelihood is only evaluated at the interior of the central patch (in red with the PCCS2+2E position at its centre, in yellow). The RELTH statistic is then estimated using the pixels of the red region, leaving a border of four pixels. The field/patch Y and X directions, at the centre of the field, match the Galactic coordinate lines of constant latitude and longitude, respectively. Each individual pixel (not drawn) is ≈ 1.′72 × 1.′72.

simulations (see Appendix B.2) to minimize the bias between the object's injected and recovered parameters. The distribution of the thermal parameters (T, β) is particularly sensitive to the value of ϕ. A high ϕ value (>50%) generates significant positive bias in β and negative bias in T, while a low value (<10%) has the opposite effect. The value we have selected (10%) leads to the lowest global bias in the main recovered source parameters, β, T, and flux density (see Appendix A.1.4).

We further assume that the *Planck* and IRIS background maps are uncorrelated, because the introduction of IRIS results in instabilities in the estimation of the covariance matrix, particularly in regions with a very bright background or with visible artefacts in IRIS.

Having fixed the cross-correlation parameters, we can proceed to run BeeP's likelihood. We first try to find the posterior maximum inside a square of at most 7 × 7 pixels centred on the original PCCS2+2E position[37]. It is often the case that the maximum of the posterior does not match the central patch pixel or that we cannot even find a maximum (see e.g., Fig. 2). If there is more than one likelihood maximum inside the search region, we always prefer the one closest to the original PCCS2+2E coordinates. It is useful to note that if a posterior maximum is not found, there is no guarantee that the derived parameter estimates follow the statistical properties predicted in Sect. 6. Whether a maximum is found or not is reported in the catalogue field MAXFOUND.

Table A.1. Relation between the instrument beam FWHM and those used in BeeP's likelihood.

Channel	Instrument $FWHM$	Likelihood $FWHM$
3000 GHz	4.′3	1.′72
857 GHz	4.′64	2.′44
545 GHz	4.′83	2.′79
353 GHz	4.′94	2.′98

The source extension parameter (EXT) poses further difficulties to an unbiased recovery of the object parameters (in this case its size). In the current implementation of BeeP, we have fixed the beam size at each frequency to an average value for the entire catalogue. According to this data model, the narrowest feature in the maps must at least have the width of the beam at that channel. However, in the real maps narrower compact objects may be present[38]. These cases create regions of the likelihood manifold with a high concentration of probability (they contain the likelihood peak) that cannot be explored because our source model does not consider "negative" radii. As a result, strong deviations in the recovered parameters for these sources can be expected. To tackle this problem we take advantage of degeneracy between the source and the beam size: a pixel brightness pattern can be the result of a narrow source and a large beam or of the reverse situation. Our solution consists of implementing simulated beams that are narrower than the average of the real beam, i.e., their FWHM is selected such that a source with an estimated size of EXT = 0.975 pixel (≈1.′72) will result in an object on the map that has the same extension as the average (real) beam (see Table A.1). This trick is actually quite important for recovering a flux-density-unbiased sample: as we can see in Fig. 8, there is a positive correlation between the source extension EXT and the flux density SREF, which implies that a bias in the estimate of EXT will propagate into SREF. Narrowing the beam artificially removes most of this bias. However, in those regions where a feature in the map is narrower than the beam size, the "source size" recovered by BeeP (EXT $\lesssim$ 1.′72) is poorly determined, since it is degenerate with the beam width. In the BeeP catalogue, we report both EXT and a more realistic source size under the field label R[39].

Once BeeP has found the parameters that maximize the likelihood, the field described in Eq. (A.13) is generated and the RELTH quantile is evaluated over the red area in Fig. A.3. We leave a 4 pixel-wide border to avoid fake likelihood maxima resulting from edge effects[40], which would artificially increase the background non-Gaussianity.

[37] As described in Sect. 3.1, we allow the optimal source location to vary from the original PCCS2+2E location by up to 3 pixels. However, some source positions may end up at a distance of slightly more than 3 pixels – this happens because when the maximum reaches the 3 pixel boundary, we allow the sampler to explore the region around the boundary.

[38] This might happen because in some regions of the sky the real beam size is narrower than the average value, or because background or noise fluctuations may cause a beam-sized object in the map to be artificially narrowed.

[39] For all cases where EXT < 1.′72, we set R = 0.0.

[40] "Edge" effects are the result, in the Fourier transform, of the discontinuities at the borders of the patch in which we are computing the likelihood. The Fourier transform requires that the data must be periodic.

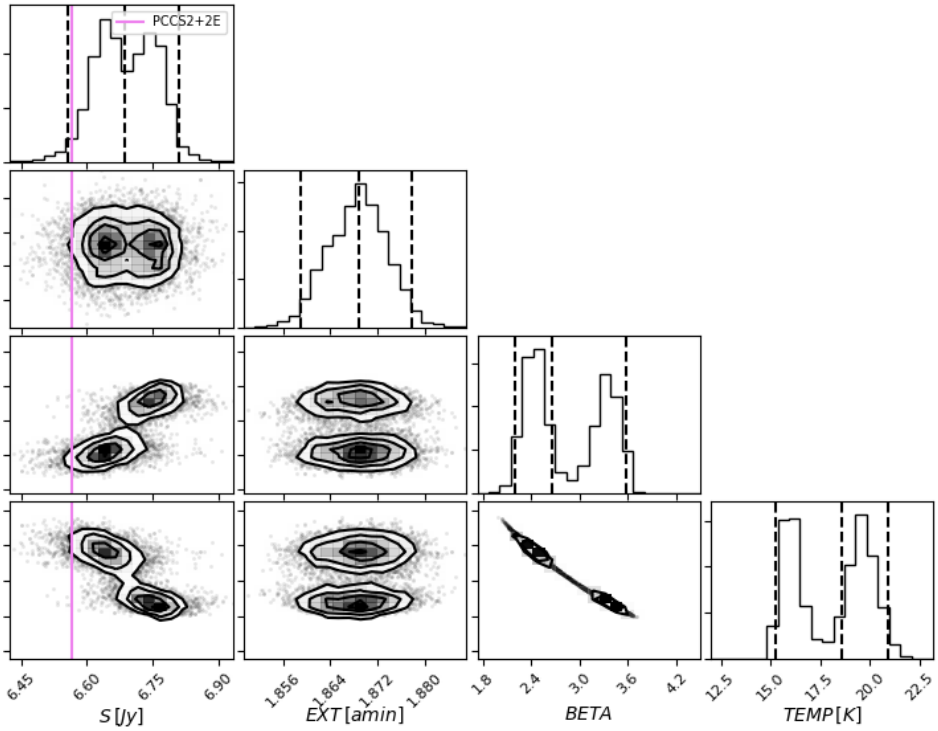

Fig. A.4. Multi-parameter fits for a particular source. This fragment of a "corner" plot (see Fig. 8) shows the multi-modal character of some of the posterior distributions.

As described in Sect. 6.2.1, we implement two different "SED models", and each of these requires a separate run of BeeP. For the secondary method (which does not impose an SED correlating frequencies), we assume that the backgrounds are independent and set the "cross-correlation factor" to 0. We use BeeP to estimate the flux density and its uncertainty in each channel independently. Then we fit an MBB curve to the individual channel flux densities using a Gaussian likelihood with variances estimated in the previous step.

A.2.4. Sampling the posterior: multi-modality

Unfortunately, owing to the very complex backgrounds, especially close to localized features like the cirrus, some posterior distributions are multi-modal (see Fig. A.4) Although there are specialized samplers that can handle multi-modal distributions like (that described in Feroz & Hobson 2008), they are far too slow for this problem. It is recommended by the authors of `MCMC Hammer` that the efficiency of the sampler can be increased by starting it in the neighbourhood of the posterior mode. We have opted for running the likelihood maximizer multiple (10) times, with initial points scattered across the prior volume, and then split the initial samples ("walkers") across the different maxima. Using this strategy we have almost completely removed the chance of the sampler systematically missing significant parts of the likelihood manifold[41]. However there are still a few cases where this solution is not effective. Some of the outliers (see Eq. (15)), in particular those with a tiny extension (smaller than the beam) or vanishingly small error bars, are the result of the MCMC chains being attracted to strong and very narrow maxima. After being caught inside these narrow local maxima, the chains are *not* able to explore the entirety of likelihood manifold, and cannot properly account for the correct parameter uncertainty, or find the global maximum. Another problem is the "chain correlation length" and dependence on the sampler "initial conditions". The first problem can be solved by periodically throwing away samples. However, one of the reason why we have chosen `MCMC Hammer` is its small correlation length. We always monitor the correlation length of the chain and when it is higher than the required level we reset the sampler and restart all chains again. The samples acceptance rate always remains very close to the optimal range of 20–50% (Goodman & Weare 2010), except for very rare occasions when it could be as low as 6%, but never higher than 62%[42]. The sample acceptance rate is reported in the ACCEPT field of the output catalogue. Perhaps surprisingly, after the first runs of the sampler (with very simple examples), we realised that the quality of the generated samples was very dependent on the sampler initial state. To overcome this difficulty we massively increased the "burn in" phase and the problem was solved. We are currently using 5000 "walkers", 98 "burn in" iterations and we only keep the two final ones, generating 10 000 posterior samples.

A.2.5. Colour correction

One important advantage of using a multi-channel estimation algorithm is that the effect of the detector finite band-passes (or "colour-correction"), can be included in the estimation chain. Although the colour correction is a relatively small adjustment[43],

[41] In the first runs of the algorithm, we were finding, in the T–β plane, an unexpected high Galactic latitude source population around 20 K. This "anomalous" population simply vanished when we adapted the code to account for multi-modal likelihoods. At the same time we also saw an increase in the parameters uncertainty, which supports the fact that we are now exploring a wider likelihood manifold.

[42] Only 151 BeeP/base catalogue sources (0.6%) have acceptance rates below 20%.

[43] Colour-correction coefficients are of the order a few percent ($\lesssim$10%). For extreme values of T ($\gtrsim$30 K) and β ($\gtrsim$3.0) they can reach values in excess of $\gtrsim$20%, but only for the 353 and 545 GHz channels.

it can introduce a bias in the MBB T–β estimates if not properly accounted for. We created a 2-d colour-correction matrix with one axis assigned to "β" and the other "T" based on the code described in Planck Collaboration IX (2014) for each *Planck* channel at 353–857 GHz, as well as IRIS. The actual correction coefficient is then obtained using bilinear interpolation.

Appendix B: Simulation-based tests

In this appendix we describe the different simulations used in this paper.

B.1. Synthetic background

This type of simulation tries to recreate a data set that follows our data assumptions as closely as possible. It is meant to verify the algorithm and code correctness under ideal circumstances. The outcome serves as a yardstick to assess the robustness of the code/algorithm as we move to more realistic cases where some of these assumptions need to be relaxed.

The diffuse background in these simulations is intended to be as close to a homogeneous Gaussian random process as possible, but with realistic *Planck* levels and characteristics. The simulations were generated from one *Planck* 2015 CMB simulation and four different noise realizations, taken from the *Planck* Legacy Archive. The CMB+noise maps were scaled in amplitude to match the median level found in each of the four real maps (*Planck* 2015 345, 545, 857 GHz, and IRIS 3000 GHz). This process ensured that the maps have signal amplitudes similar to those found in the real maps, but their statistical properties are Gaussian. Then we cut the spherical maps into many small patches and we injected a source directly into the centre of each patch. All injected sources were simulated to be equally shaped, following a bi-dimensional symmetrical Gaussian profile with constant and very small radius, and then convolved with the PSF at each frequency, which was assumed to be constant and equal to *Planck*'s average effective beam. The sources were rendered in very high resolution and projected directly into the patch pixels. The source SEDs were derived from an MBB law with three free parameters: T, temperature; β, spectral index; and S_{857}, flux density at 857 GHz. The values for the source SED parameters were then randomly drawn from a preliminary catalogue that had been extracted with BeeP from the real maps. That precursor catalogue was cross-matched with the PCCS2+2E and the GCC. The parameter estimates showed a high degree of consistency with both catalogues and were thereafter assumed as representative of the actual sky distribution. The goal of this type of simulation is to closely replicate the assumptions of our data model, and therefore constitute our "fiducial case".

In Fig. B.1 we make a comparison between the reference flux density of the injected sources with those we retrieved using BeeP. Owing to the huge dynamic range of the values (five orders of magnitude), computing the distribution of the fractional difference gives a better understanding of their consistency than just the difference. As explained in Sect. 6.2.4, instead of plotting the fractional difference formula directly ((out − in)/in), we replace it with ln(out/in). Figure B.1 depicts two flux retrieval cases, with two different patch sizes, to gauge its impact on the recovery precision (see Fig. A.1.5). The top panel shows the comparison when the "fields" cut from the homogeneous background sphere were 513 × 513 pixels and the "patches" (core region where the likelihood is evaluated) 129 × 129 pixels. The middle panel of Fig. B.1 shows exactly the same thing, but the dimensions of the fields were this time

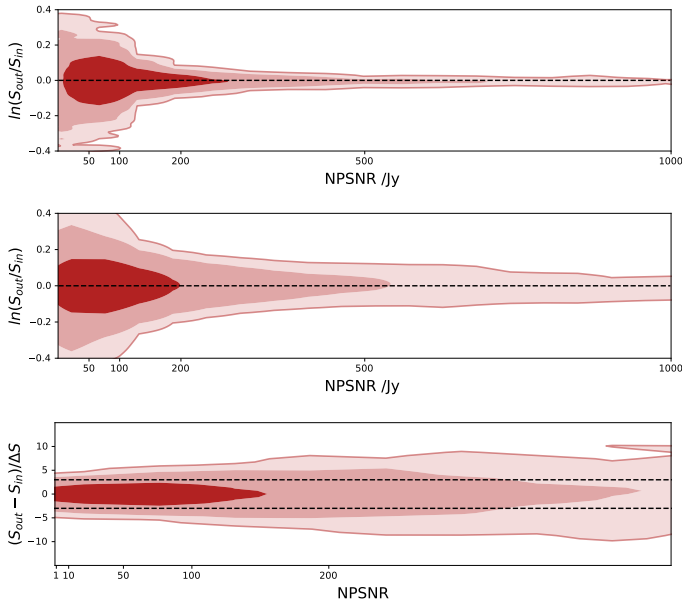

Fig. B.1. Comparison of input and output flux densities. *Upper* and *middle panels*: $\ln(S_{\rm out}/S_{\rm in})$ versus NPSNR distribution contours ([68, 95, 99]%). In the *upper panel* the field and patch are 513 × 513 and 129 × 129 pixels, respectively, and in the *middle* 129 × 129 and 33 × 33. For both cases a source was directly injected in the central pixel of the patch, but always with a small random shift from the pixel centre. *Lower panel*: "normalized error" = $(S_{\rm out} - S_{\rm in})/\Delta S$ distribution contours for the small patches. The horizontal lines in the bottom panel, are the $\pm 3\,\sigma$ boundaries.

129 × 129 and 33 × 33 pixels. The retrieved flux distributions are similar and both show that after a certain NPSNR threshold the precision of the estimates saturates. However, the top panel of Fig. B.1 (larger field and patch) shows much less dispersion, especially as we move towards higher NPSNR values, and it reaches saturation much later. This should not have come as a surprise. As was mentioned in Appendix A.1.5, in the high S/N regime, the uncertainty in the parameters recovery is limited by the estimation accuracy of the covariance matrix. The larger the data set, the more precise the estimation is, the less dispersion the estimates show, the later the onset of the flux accuracy saturation. However, when tackling the "real world", the smaller the patches the better the background homogeneity assumption actually holds. From our simulation exercises, fields of 129 × 129 and patches of 33 × 33 pixels seem to provide the best balance between background homogeneity and enough data (i.e., pixels) to guarantee that the error in the statistics we collect do not dominate (for the majority of cases). However, when assuming that the covariance matrix had no estimation error, we failed to propagate into the likelihood that extra source of uncertainty arising from the field/patch statistics. That will necessarily lead to an underestimation of the error bars. For the low NPSNR regime this is not a problem because the covariance-matrix estimation error is still sub-dominant; however, at the high end where it completely dominates, the error bars are underestimated (Fig. B.1 lower panel).

Since in practical terms it is impossible to propagate the covariance-matrix uncertainty into the likelihood (see Appendix A.1.5), we have chosen to keep the likelihood as it is, but later correct the error bars for the unaccounted uncertainty.

Another source of systematic errors could stem from the projection of the compact objects onto the flat fields. To study the

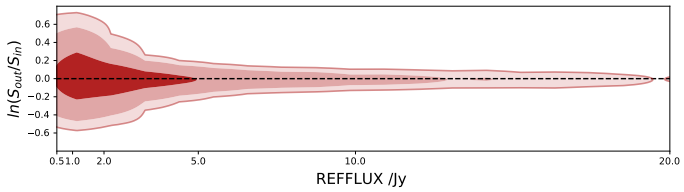

Fig. B.2. Recovery error versus source brightness. This specifically shows the $\ln(S_{\rm out}/S_{\rm in})$ versus $S_{\rm in}$ distribution contours ([68, 95, 99]%). The sources were injected into the CMB+noise only HEALPix maps following the same process as when injecting into the real data. The same source was injected multiple times to assess the impact of the different background conditions.

effect of a potential projection distortion, we also injected the sources directly into *Planck*'s CMB + noise simulated maps at *Planck*'s HFI native resolution (HEALPix $N_{\rm side} = 2048$). A minimum distance (12′) between injected sources was imposed in order to avoid beam blending effects. Figure B.2 shows once again the quantity $\ln(S_{\rm out}/S_{\rm in})$ but this time versus $S_{\rm in}$. We injected the same source multiple times to learn how the different background conditions would affect the extraction. The background is homogeneous, but its dominant component (the scaled CMB) has a typical correlation length of around 1°, which is much larger than the typical extension of a source, around 10′. This implies that some of the injected sources sit on top of (scaled CMB) crests and others on valleys.

B.2. Injecting simulated sources into real maps

The sky distribution of the PCCS2+2E, our input catalogue, is extremely inhomogeneous. The PCCS2 contains less than 10% of the sources and covers about half of the sky. The PCCS2E (≳90%) sources are located almost entirely within regions of strong, complex background emission. Planck Collaboration XXVI (2016) indicates that a realistic rendering of the cirrus, including localized embedded features, is absolutely crucial for validating any catalogue at the frequencies considered here. Unfortunately, simulating the submillimetre sky is a formidable task (Delabrouille et al. 2013), and realistic simulations of such a complex background are not yet available. Since BeeP's model of the background is a statistical one, it is critical that the statistical properties of the simulated background match those of the real data. To achieve this match, we used the actual *Planck* 2015 maps and injected fake sources directly into them. This approach is similar to the one previously employed in the production of *Planck*'s Early Release Compact Source Catalogue (Planck Collaboration VII 2011). The physical parameters of the mock sources correspond to those of the original sources, as extracted from a preliminary run of BeeP on the 2015 maps.

Given the complexity of the Galactic background, the mock sources should ideally be injected exactly on top of the PCCS2+2E catalogue positions. However, this is only possible if the real source is first removed in such a way that the background where it is to be embedded is left undisturbed; otherwise residuals of the removed source could systematically bias the extraction results. To try to accomplish this, we mask the pixels around each real source and then inpaint them by diffusing the background into the masked region, starting from its boundaries; the inpainting method is described in Casaponsa et al. (2013) and preserves the statistical properties of the field surrounding the inpainted area. The radius of the inpainting mask is 7′, and we impose the condition that the minimum distance between any two injected sources should never be smaller than 12′. As a consequence, some source positions in the PCCS2+2E do not have any source injected, a situation that happens more frequently at low Galactic latitudes.

We recognize the possible bias of injecting sources into a modified sky. Therefore, to validate the inpainting procedure we also generate a second set of simulations in which each mock source is injected not at the original PCCS2+2E location but in its near neighbourhood. For this set of injections, we place a mock source within an annulus around the original position, within a radius of 12′ and outside a radius of 20′[44]; we also ensure that the injected source does not blend with any other source previously injected or in the PCCS2+2E. This mechanism guarantees that no source is ever injected in an inpainted area. Because of this restriction, in regions of a very high source density, such as the Galactic plane, there may be some PCCS2+2E source locations that are not associated with any injected source. We note that for these simulations we must also inpaint the original source location, otherwise it would *systematically* increase the non-Gaussianity of the background patch under analysis. Sources of equivalent flux densities would *always* appear in pairs, making the original PCCS2+2E source systematically increase the background non-Gaussianity of the injected source background. The annulus, however, lies well outside the inpainted region and guarantees (given the equivalent beam width) that the condition of Eq. (A.9) always applies and that any background disturbance, such as another source or an inpainted hole, does not perturb the parameter estimation. As may be seen below, both types of simulations produce statistically similar sets of results (see Table B.1).

For the sake of completeness, we also add a third set of simulations in which the injected mock sources are uniformly distributed on the sky; as in the previous simulations we make sure that sources do not overlap with any other source in the PCCS2+2E. For this case, we draw the mock source parameters at random from the PCCS2 sub-catalogue rather than from the full PCCS2+2E (otherwise, we would create an unrealistically bright high-Galactic-latitude population that would systematically increase the catalogue source significance.

The three types of simulations (just described), were employed to calibrate the "cross-correlation factor" (ϕ) (see Appendix A.1.4). We applied BeeP using a set of ϕ values, {0.0, 0.05, 0.10, 0.15, 0.20, 0.30, 0.50}. We then looked for bias by comparing recovered to injected source properties. The simulations showed that the value of ϕ used has a strong impact on the recovery of the MBB parameters (T, β), and must be chosen with care. What we learned from this exercise[45] is that faint background regions prefer higher ($\phi \approx 0.15$) values and complex, bright regions lower values ($\phi \approx 0.05$). So we selected $\phi = 0.10$ as a balance between the two cases. In any event, the bias in T and β that was observed when ϕ was not optimal for a particular subset was always small (≲5%).

[44] For this case we implement a wider range of distances between the original and the injected source than in the "no-sources" simulation, to avoid the risk of systematically creating pairs of identical sources, but retaining a similar background as that of the original source.

[45] When $\phi \geq 0.20$ all three sets displayed negative bias in T and a positive in β. For $\phi = 0$ all three cases showed exactly the opposite trend. The value of ϕ that minimized the bias for the "uniform" injection policy was 0.15. For the "in place" and "neighbourhood" policies the best ϕ was 0.05. However, if we selected only the PCCS2 subset for these two cases, the optimal value of ϕ would return to 0.15.

With ϕ set, we now examine the accuracy of the recovery by BeeP of the physical source parameters in the simulations. For this purpose, we define now the "normalized symmetric error" (Δ) of a parameter as

$$\Delta = \frac{4(\theta_{\rm out} - \theta_{\rm in})}{\epsilon_{\rm h2\sigma} - \epsilon_{\rm l2\sigma}}, \tag{B.1}$$

where $\theta_{\rm in}$ is the injected parameter value, $\theta_{\rm out}$ is BeeP's best estimate and $\epsilon_{\rm h2\sigma}$, $\epsilon_{\rm l2\sigma}$ are the 97.5% and 2.5% values of the distribution, respectively. In the case of an ideal Gaussian distribution of errors, the variable Δ should follow a unitary normal distribution. However, because of the strong inhomogeneity and non-Gaussianity of the background, especially at low Galactic latitudes, it is expected that sub-optimal parameters will be found for certain objects, i.e., that in some parameter ranges significant outliers will arise in the distribution of errors. As one of our goals, perhaps the most important, is to have a well defined set of statistical traits for the catalogue estimates, these extreme outliers need to be identified and removed to avoid biasing or distorting the characterization.

The set of simulations previously described was used to identify such cases, which appear in the results as sources with either unphysical parameters or vanishingly small parameter uncertainties. Both manifestations are signs of poor or insufficient posterior sampling, raising the possibility that entire ranges of feasible parameter values are not being sampled by the likelihood exploration tool (for instance when the likelihood is multimodal). When this happens, one cannot be sure that the optimal estimate set was obtained and the error bars are certainly underestimated[46]. We are now able to define exclusion regions in parameter space. We have found that the vast majority of sources whose estimated parameters do *not* meet the following criterion should be flagged:

$$\begin{aligned} &\mathsf{EXT} > 1.46 \wedge \mathsf{TEMP} < 60 \wedge \mathsf{BETA} < 5 \\ &\quad \wedge (\mathsf{TH2SB} - \mathsf{TL2SB}) > 0.8 \\ &\quad \wedge (\mathsf{BETAH2SB} - \mathsf{BETAL2SB}) > 0.25. \end{aligned} \tag{B.2}$$

Such sources are severe outliers in one or more parameters[47], and therefore should be discarded. This is the "outliers criterion"[48]. In Fig. B.3 we show the distribution of Δ (Eq. (B.1)) for the MBB parameters TEMP and BETA, as well as for SREF. In this figure we show the "same-location" simulations on the left and the "neighbourhood" simulations on the right. We did not include in this assessment a small fraction of simulated objects that behaved anomalously for other (understood) reasons[49].

[46] The catalogue also contains a field ACCEPT with the sampler acceptance rate. Very low (<10%) or very high (>80%) values are signs of a sub-optimal likelihood exploration. For further details see Appendix A.2.4.

[47] Δ_β and Δ_T at least > 7.

[48] In the BeeP/base catalogue, only 999 of the sources that pass the "reliability criterion" (Eq. (B.2); about 2% of the entire PCCS2+2E) are rejected by the "outliers rejection" condition.

[49] We omit in particular those whose recovered positions moved by more than 0.′8 (about half a pixel), since the estimates at these new locations cannot be directly compared to those of the injected objects; in fact this is only a tiny fraction of the simulated catalogues (50 sources in the "same-location" simulations). In addition, we have noted that some source locations, especially around extended objects, do not coincide with the actual centre of the object; the removal and inpainting process are not effective in these cases (374 sources). Finally, close to the Galactic plane where the PCCS2E is hardly complete, in some instances BeeP prefers the location of a nearby object that was not in the PCCS2E. Very rarely, this also occurs when the injected source is extremely faint.

Table B.1. Statistics of the MBB $\{T,\beta\}$ and SREF recovered parameters.

Data set	Parameter	$\langle\Delta\rangle$	Median $\ln(\theta_{\rm out}/\theta_{\rm in})$	σ_Δ	SMAD [(a)]
		Injection in neighbourhood			
PCCS2	TEMP	−0.01	1.5%	0.74	0.64
	BETA	−0.19	−6.4%	0.83	0.68
	SREF	0.30	4.0%	1.63	1.36
PCSS2+2E	TEMP	−0.33	−3.2%	0.77	0.68
	BETA	0.13	3.7%	0.74	0.62
	SREF	0.35	6.3%	1.72	1.38
		Injection at same location			
PCCS2	TEMP	−0.03	0.8%	0.69	0.59
	BETA	−0.15	−5.4%	0.75	0.61
	SREF	−0.04	0.2%	1.35	1.00
PCCS2+2E	TEMP	−0.38	−3.3%	0.80	0.60
	BETA	0.16	3.2%	0.70	0.55
	SREF	0.40	6.5%	1.66	1.12

Notes. The data sets were filtered with the intersection of the "reliability" and "outliers rejection" criteria. [(a)]Scaled median absolute deviation.

Table B.1 shows a statistical summary of the offsets in parameters shown in Fig. B.3. To reduce sensitivity to the presence of outliers, we have replaced the usual "average" and "standard deviation" with the more robust "median" and "scaled median absolute deviation" (SMAD).

For both types of simulation, Table B.1 shows that BeeP recovers T and β in a largely unbiased manner. In addition, Fig. B.3 shows that the uncertainties in T and β are only slightly overestimated. They also show a significant correlation (see Fig. B.4). This is not unexpected: inspecting the posteriors for individual sources (see e.g., Fig. 8) we see a strong banana-shaped degeneracy between these two parameters. The flux-density recovery statistics depict a slightly different situation. We note that in these simulations the injected sources are circularly symmetric and beam shaped, in accord with the data model of BeeP. Even in this benign situation, inspection of Fig. B.3 and Table B.1 indicates that the dispersion of the flux-density estimates is larger than expected. We must conclude that BeeP underestimates the uncertainty of the recovered values of SREF. On the other hand, for the "same location" simulations, those that should best reproduce the real extraction conditions, in the PCCS2 region, the SMAD statistic shows values equal to or below 1, even for SREF. When extending to the full PCCS2+2E only a small excess appears. Based on this we could conclude that, unless a rigorous statistical characterization of the estimates is necessary, the uncertainty values as given in the catalogue, are fit for the purpose.

At the same time, we know that these simulations are not realistic enough to provide a proper assessment of the retrieval of the flux density. Therefore, we postpone the discussion of flux-density recovery bias to the next sections (see Appendices B.3 and B.4).

These simulations also allow us to examine the quality of recovery of the source locations. Figure B.5 shows histograms of the separation between the injected and estimated source position (SRCSEP). When injecting at the same position as the PCCS2+2E (in blue) we find a small bias (around 0.′1–0.′2). If now we normalise SRCSEP with POSERR (right panel), we find that the "same place" simulation overerestimates the

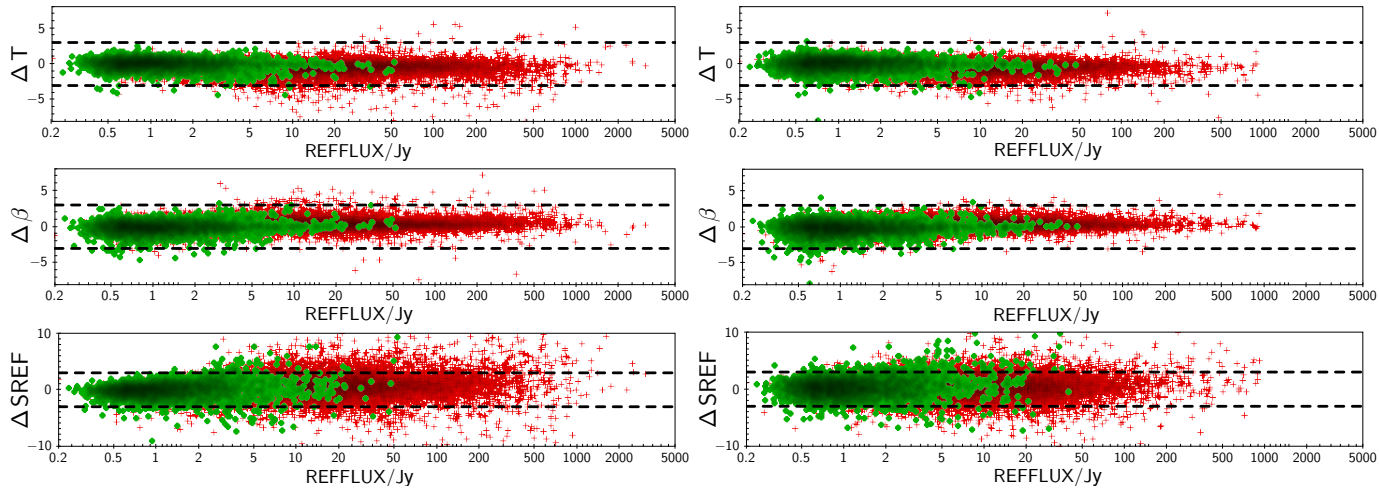

Fig. B.3. Errors in simulated source properties. The y-axis shows the normalized symmetric error as defined in Eq. (B.1) (Δ) and the x-axis the injected source flux density (REFFLUX). The data set used here is the intersection of the reliability and outliers criteria. The red crosses show all sources, while the green circles are just the PCCS2. The horizontal dashed lines are at $y = \{-3, +3\}$. The *panels on the left* were drawn when the simulated sources were injected in the same positions as the PCCS2+2E and those on the *right* for sources injected in their neighbourhood. Those sources whose recovered position moved by more than 0.′8 (about half a pixel) from the injection location were removed from the set. The total number filtered out was about 0.02% (50 sources) for the case of injection at the same location; for injection in the neighbourhood, the total number filtered out rose to 1.63% (374 sources). The close similarity between the two sets of plots shows that the effect of inpainting is small. Table B.1 contains a statistical analysis summary of these results.

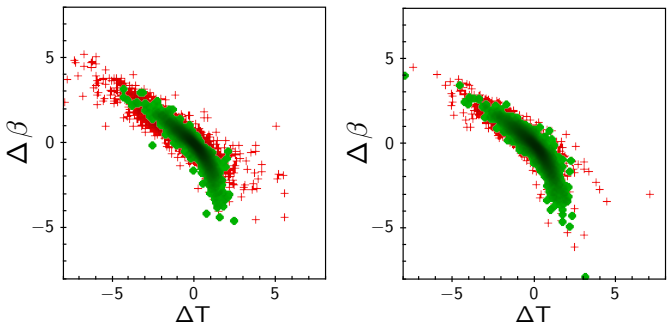

Fig. B.4. Correlation between $\Delta\beta$ and ΔT. This uses the same data set and symbols as in Fig. B.3. *Left panel*: correlation for the "same position" injection criterion and *right panel*: for the "neighbourhood" criterion.

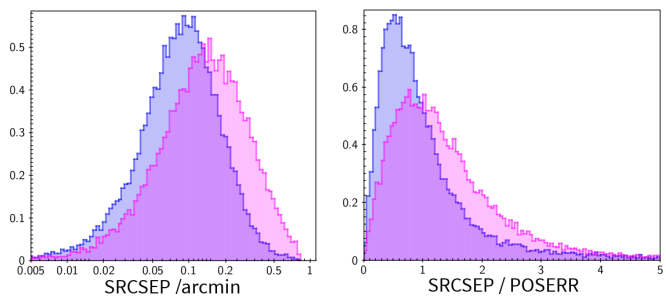

Fig. B.5. Histograms of the separation between the injected and estimated source positions (SRCSEP). *Left panel*: absolute deviation and *right*: deviation normalized by the position error bar (POSERR). The results of the "same place" simulation are in blue and the "neighbourhood" ones in pink.

position error (since we expect the normalized distribution to peak at 1). On the other hand, the histogram of the normalized separation for the "neighbourhood" shows the expected statistical behaviour[50]. We believe that this might be the result of the inpainting of the original source. It shows the usefulness of the "neighbourhood" simulation, and since its results are closer to expectation than the "same place" simulation, we use it in what follows.

In Fig. B.6 (upper left panel), the blue-purple set shows the dependence of the distance between the injection and recovered position normalized by the estimated uncertainty (POSERR), as a function of NPSNR. In this figure it would appear that there are a number of cases where the location is severely misestimated (those above the dashed line); when we restrict the catalogue to the PCCS2 set, we see that these cases correspond preferentially to high NPSNR values. In the lower left panel we can see a very strong correlation between the catalogue POSERR (blue and purple) and NPSNR[51]. We note that the estimated positional uncertainty for sources with NPSNR> 20 is very small (<1.5% of the beam size). It seems clear that the anomalous cases in the upper-left figure are mainly due to a serious underestimation of the positional uncertainty for sources with high NPSNR. The most likely reason for this is outlined in Appendix A.1.5. In fact, a similar situation was found when producing the PCCS2, and it was handled by adding a term (see Eq. (7) in Planck Collaboration XXVI 2016) that forced the positional uncertainty to remain above a threshold. We follow suit by adding a term σ_0 to our estimate of the full positional uncertainty:

$$\sigma_c^2 = \mathrm{POSERR}^2 + \sigma_0^2, \tag{B.3}$$

where σ_c is the corrected position error bar and σ_0 the saturation constant. For high NPSNR, POSERR $\rightarrow 0$ and $\sigma_c \approx \sigma_0$. In order to determine σ_0, we created a likelihood based on the

[50] The normalized position deviation should follow a unitary Rayleigh distribution.

[51] The PCCS2 set (blue points) POSERR dependence on NPSNR is well modelled by POSERR $\approx \alpha\,\mathrm{NPSNR}^{-1.01}$, with $r = -0.98$, where α is an arbitrary proportionality constant.

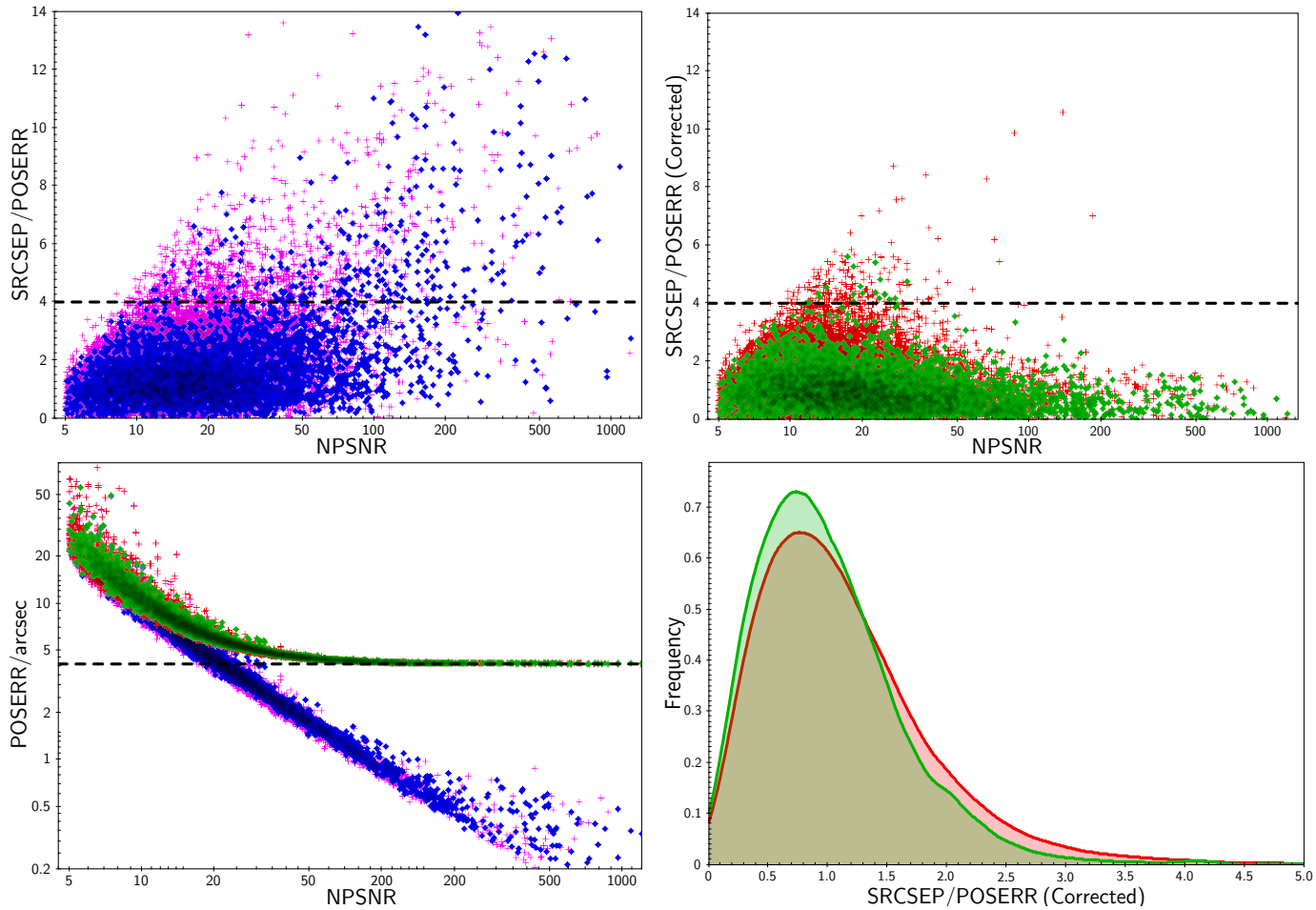

Fig. B.6. Position-recovery accuracy for a simulation with injection in the "neighbourhood" of PCCS2+2E source locations. *Upper left panel*: distance between the injection and recovered position, normalized by the estimated uncertainty (POSERR), as a function of NPSNR. The blue points represent the PCCS2 and the magenta the full PCCS2+2E. *Upper right panel*: same but after applying to POSERR the correction suggested in Eq. (B.3) with $\sigma_0 = 4\farcs11$ (here green is PCCS2 and red is PCCS2+2E). *Lower left panel*: POSERR for both sets, corrected and uncorrected, as a function of NPSNR. The horizontal dashed line is the positional precision saturation constant σ_0. *Lower right panel*: histogram of the separation between the injected and estimated source position, normalized by the corrected position error bar. The distribution is consistent with a unitary Rayleigh distributions with a minor excess in the tail.

Rayleigh distribution and we sampled from σ_0, to find

$$\sigma_0 = 4\farcs11 \pm 0\farcs03. \tag{B.4}$$

In Fig. B.6 (upper right and lower left panels), we show the corrected position error σ_c as a function of NPSNR (PCCS2+2E in red and PCCS2 in green). The horizontal dashed line is σ_0. We can see that above NPSNR $\gtrsim 20$ the positional uncertainty stops reducing and instead saturates at σ_0. In the lower right panel we show the histograms of the normalized position distribution but now using σ_c. The PCCS2 distribution (green) is now a good match to a Rayleigh distribution[52]. The PCCS2+2E (red) is also a good match, though it has a small excess in the tail. This is the same excess seen in the vertical direction in the upper right panel. For this simulation, the median position-corrected error bar is $7\farcs8$ or 7.6% of the pixel size.

We stress that the vast majority of sources with NPNSR > 20 have a well-determined positional uncertainty without any correction – i.e., those well below the dashed line in Fig. B.6 (upper left panel); applying the correction penalizes those sources unnecessarily. For this reason, the correction on POSERR described here is *not* applied to the output of BeeP, and should be used only for statistical characterization of samples of sources that contain high NPNSR sources.

B.3. Flux density uncertainty correction due to source extension

In every simulation we have described so far, the injected sources have always been beam shaped. However, as can be seen in Fig. 9, PCCS2+2E contains many sources that are at least slightly extended.

In order to address the effects of source extension, we create a new set of simulations, following exactly the same procedure as described in Appendix B.2, i.e., including three types of simulation, each following one of our injection policies. The only difference is that this time we inject extended rather than point sources. The source size parameter is sampled from a preliminary run of BeeP on the real data.

Figure B.7 (top row) shows, in blue contours, the normalized difference between the injected and recovered flux densities (Eq. (B.1)) as a function of NPSNR. In the high NPSNR regime all three types show an excess of deviations ($\gtrsim 3\,\sigma$), reflecting

[52] If σ_c is an accurate description of the actual position errors, then the normalized position distribution should follow a Rayleigh distribution with a scale parameter equal to σ_c.

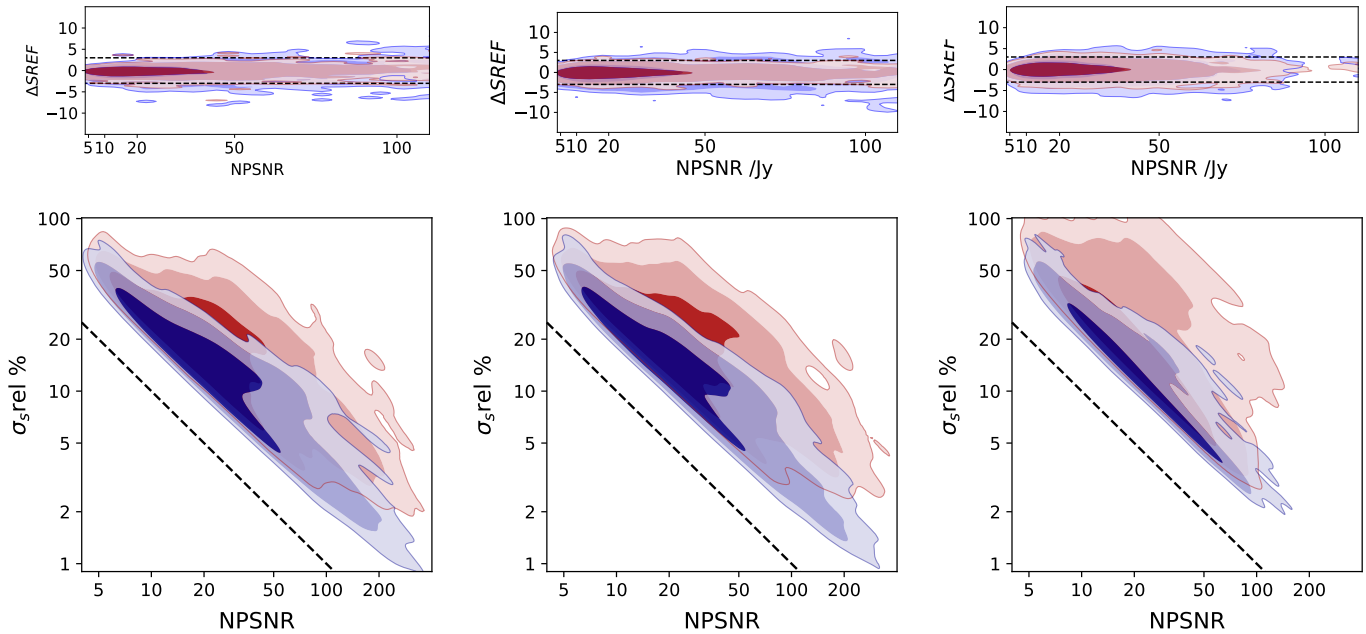

Fig. B.7. Comparison of injected and recovered flux densities. *Upper row*: distribution of the normalized difference (Eq. (B.1), blue contours, [68, 95, 99]%) versus NPSNR for three different data sets (*from left to right*): PCCS2 "same location"; PCCS2 "neighbourhood"; and PCCS2E $\wedge\,|b| > 10°$ "neighbourhood". The red contours show the same as the blue ones, but this time for corrected values (Eq. (B.5)). The horizontal dashed lines are $\pm 3\,\sigma$. *Lower row*: identical flux-density relative error (ΔSREF/SREF) distribution contours. The black dashed lines here are NPSNR^{-1}, the lower limit of the flux-density relative error, which is only achievable if the only unknown parameter in the model is the flux density.

the fact that their distribution is not fully Gaussian. This is not unexpected, since the same behaviour has been seen for the "fiducial" case simulations (see Appendix B.1 and Fig. B.1), and the same mechanism should be at work here[53]. These deviations do not affect significantly the $1\,\sigma$ levels of the uncertainty distribution, but only its tails. This non-Gaussianity should be taken into account only if it is desired to make a statistical analysis of the flux-density uncertainties of large populations of sources. To account for this effect, it is possible to add a correction, which we describe in the rest of this subsection. But we recommend that users interested in the $1\,\sigma$ uncertainty of individual sources do *not* apply this correction.

Figure B.7 suggests that a correction proportional to ln(NPSNR) would be adequate. This correction is to be added in quadrature to the error bars extracted by BeeP. For this purpose we have defined a new variable

$$\sigma_{\rm s} \equiv \sqrt{\sigma_{\rm cat}^2 + (c * \ln(\mathrm{NPSNR}))^2}, \tag{B.5}$$

where $\sigma_{\rm cat} \equiv (S_{\rm h2\sigma} - S_{\rm l2\sigma})/4$ and c is the flux-density correction constant. To compute the optimal value of c we follow a similar procedure as that for the positional accuracy. Let us define a new variable ξ as

$$\xi \equiv \frac{(S_{\rm out} - S_{\rm in}) - b}{\sigma_{\rm s}(c)}, \tag{B.6}$$

where $\sigma_{\rm s}$ is the "corrected" flux-density uncertainty and b is a "bias", which is added to help symmetrize ξ. If the BeeP uncertainty ($\sigma_{\rm cat}$) were a truthful representation of the uncertainty, in a Gaussian sense, then ξ would follow a normal distribution with $b = 0, c = 0$. We characterize a Gaussian likelihood for ξ with two parameters $\{b, c\}$. We sample from $\{b, c\}$ to construct a posterior distribution and then we find the median of both parameters to correct the catalogue. We expect that in regions with strong complex backgrounds, the sub-optimality of BeeP's likelihood

Table B.2. Flux density error bar calibration constant "c" in mJy, for the three types of simulation (see Eqs. (B.5) and (B.6)).

Data set [(a)]	c [mJy]	Before [(b)] σ_ξ [(d)]	Before [(b)] SMAD_ξ [(e)]	After [(c)] σ_ξ	After [(c)] SMAD_ξ [(e)]
	Injection at same location				
P2E > |10°| [(f)]	103 ± 2	1.38	0.95	1.19	0.81
PCCS2	48 ± 1	1.51	0.91	0.91	0.66
< |10°|	621 ± 15	1.54	0.99	1.28	0.77
	Injection in neighbourhood				
P2E > |10°|	168 ± 2	1.59	1.38	1.17	0.93
PCCS2	57 ± 1	1.70	1.31	1.00	0.88
< |10°|	820 ± 15	1.66	1.40	1.17	0.91
	Uniform distribution				
P2E > |10°|	85 ± 2	1.37	1.14	1.06	0.91
PCCS2	57 ± 1	1.38	1.21	1.07	0.93
< |10°| [(g)]	259 ± 15	1.60	1.17	1.02	0.86

Notes. [(a)]"P2E > 10°" means the PCCS2E data set with $|b| > 10°$; "PCCS2" means the PCCS2 data set; and "<|10°" means the PCCS2+2E set with $|b| < 10°$. [(b)]Before applying the correction (as in the catalogue). [(c)]After applying the correction. [(d)]Standard deviation of ξ (Eq. (B.6)). [(e)]Scaled median absolute deviation of ξ. [(f)]The median NPSNR for all subsets is approximately 20, except for "< |10°|", which is approximately 9 and "PCCS2E > |10°|", which is approximately 13. [(g)]There are only 1066 sources in this subset.

[53] The propagation of the covariance-matrix uncertainty into the likelihood results in an underestimation of the flux-density uncertainty.

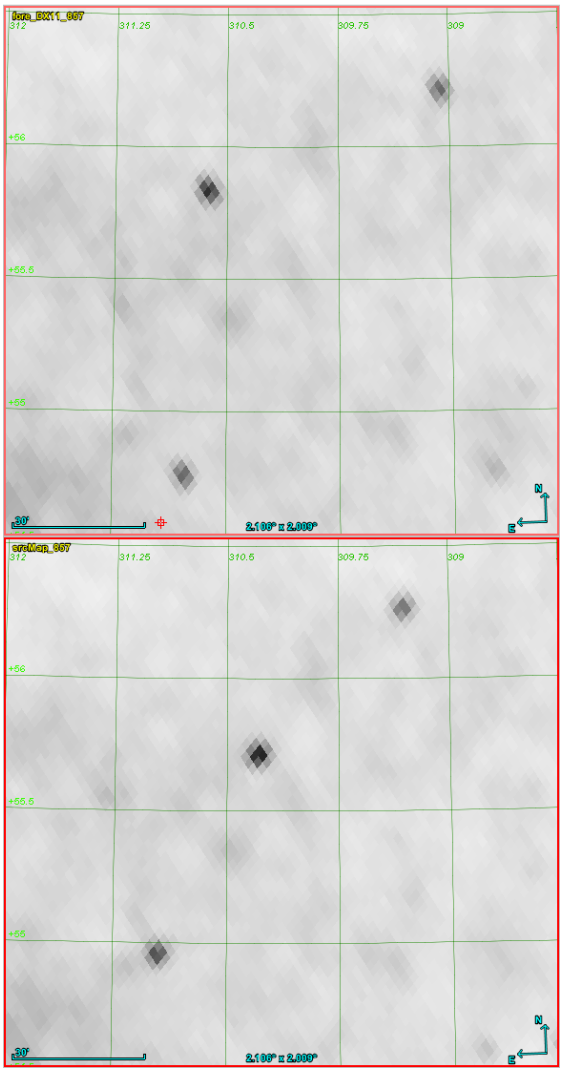

Fig. B.8. Small field ($2.^\circ1 \times 2.^\circ0$) centred at $l = 309.^\circ68, b = +55.^\circ41$. *Upper panel*: *Planck*'s 857 GHz map and *lower*: BeeP injection simulation where mock sources were added to the same *Planck* map around the neighbourhood of real sources that were masked and inpainted. The simulated sources were rendered using the average effective Gaussian 857 GHz *Planck* beam. The brightness scale is the same on both plots. One can see in the *lower panel* that the injection simulations fail to capture the ellipticity of compact objects in the *Planck* 857 GHz map (*upper panel*).

will manifest itself more strongly and require larger corrections. We therefore compute corrections for each of three sky regions:

- PCCS2;
- PCCS2E $\wedge\, |b| > 10^\circ$;
- $|b| \leq 10^\circ$.

In Fig. B.7 (lower row) we show the flux-density relative accuracy as a function of NPNSR, before (in blue contours) and after applying the correction (in red contours). The corrected error bars show (as expected) a larger dispersion with NPSNR.

Table B.2 contains a summary of statistics of the variable ξ (Eq. (B.6)) for the three type of simulations and the three sky regions before and after applying the correction. If the error bars were correctly describing the flux recovery uncertainty, in a Gaussian sense, then $\sigma_\xi \approx 1$. Considering that the data statistics are very non-Gaussian, with broad tails, and σ_ξ is very sensitive to outliers, we also included the more robust scaled median absolute deviation (SMAD$_\xi$). "Before" applying the correction,

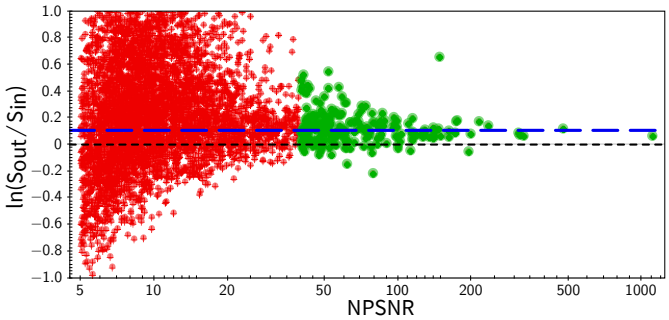

Fig. B.9. Comparison of FFP8's full catalogue of bright source flux densities with BeeP's SREF. Shown in red are the source flux estimates with NPSNR < 40 and in green those with NPSNR > 40. The horizontal long-dashed blue line is the $\ln(S_{\rm out}/S_{\rm in})$ median (0.104) of the green subset. The strong outlier is the result of a mock source blended with a real one.

all sets, for all simulations, show a clear excess in σ_ξ. This is also seen, as expected, in the SMAD for ξ, for the cases "neighbourhood" and "uniform", but not for "same location". By examining Fig. B.7 (left column, in green), one can indeed see that the distribution of estimates for "same location" is tighter than for the other types. We also find for "same location" a small bias towards low values, which is often present in this type of simulations (e.g., Fig. B.3). We believe this bias could be an effect induced by the inpainting procedure.

On the other hand, "After" correcting the flux-density error, all three types of simulations show reasonable values for both statistics, although in the case of "same location" simulations (which is perhaps the most realistic), the improvement is not as good as in the other types. However, as previously mentioned, the SMAD statistic (the more robust measurement of dispersion for non-Gaussian distributions) applied without any correction ($\{b, c\} = 0$) to the "same location" simulation already showed very good values (see Table B.2). For most purposes it should therefore not be necessary to apply any correction to the BeeP estimates of flux-density uncertainties. Corrections should be applied only if a strictly Gaussian characterization of the uncertainties is needed, particularly true for high-Galactic-latitude objects.

B.4. Planck FFP8 simulations

Possibly the crudest part of the data model implemented by BeeP is that it assumes that the beam shapes are perfectly circularly symmetric and homogeneous across the sky. In addition, in our injection simulations, the mock sources always have circularly symmetric Gaussian shapes, and in the most sophisticated simulations we also vary their radius. However, for the *Planck* 857 GHz channel, the average beam ellipticity ($\varepsilon \approx 1.39$) is sufficiently high and variable across the sky (dispersion about 10%), to induce systematic flux-density deviations as a result of the model and actual beam-shape mismatch. Given the huge flux-density dynamic range of the PCCS2+2E, we expect that, especially at the bright end, these effects will have a significant influence on the estimation of flux densities. These systematic effects cannot be directly taken into account by BeeP.

In principle BeeP's likelihood can easily accommodate more realistic beam shapes, including their spatial variation[54], but

[54] For example, as described in the FEBeCoP effective beam approach (Mitra et al. 2011).

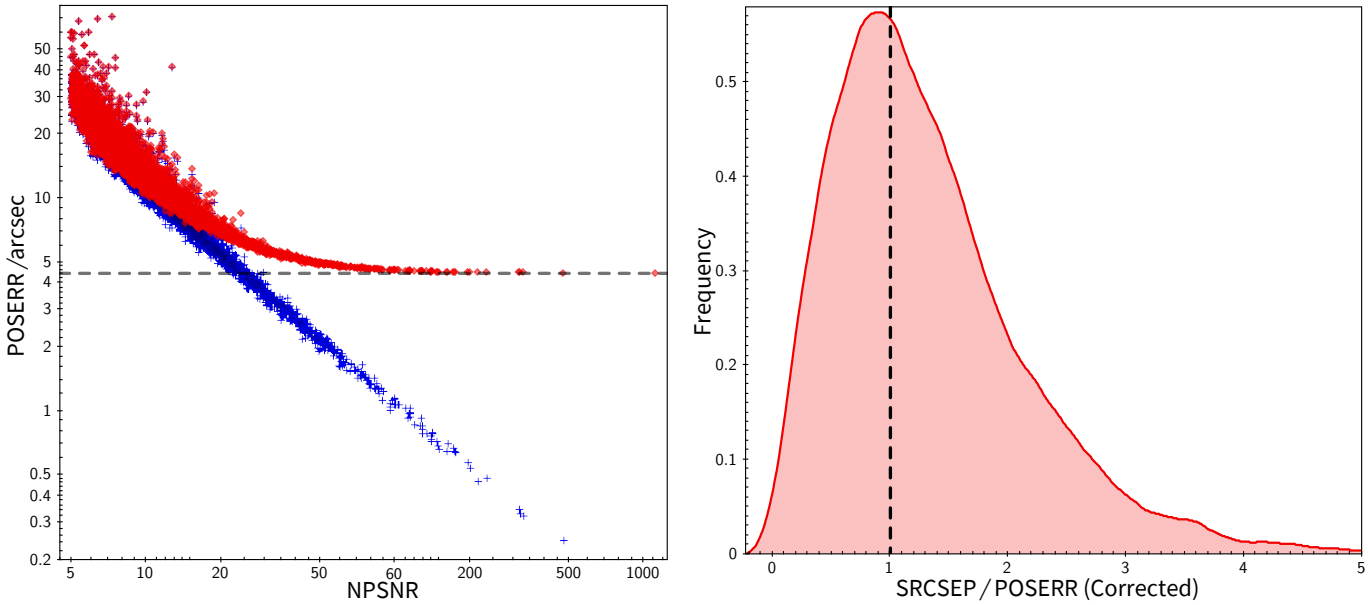

Fig. B.10. FFP8 positional accuracy results. *Left*: raw (as in the catalogue) and corrected POSERR versus NPSNR (FFP8 bright sources catalogue). The correction procedure was the same as in Appendix B.2. The horizontal line is the saturation constant σ_0. *Right*: histogram of the deviations between the injected and recovered positions, normalized by the corrected positional uncertainty.

the computational cost would be prohibitive. However, *Planck* has produced a set of simulations that include a very accurate model of the beam shapes and their variation across the sky, the "FFP8" simulations (Planck Collaboration XII 2016). We note two important drawbacks of FFP8: the absence of a 3000 GHz map; and the fact that all simulated compact objects in the maps are exactly beam shaped (they are drawn from a simulated set of zero-extension sources). These issues affect the constraining capability of BeeP. With only *Planck*'s three high-frequency channels available, BeeP can no longer effectively constrain T. The extra uncertainty propagates to β (they are highly correlated) and to a smaller extent to the flux density S (see Fig. 8). In spite of these drawbacks, we can use the FFP8 simulations to effectively calibrate the effect of beam shapes on the recovery of source parameters.

We are particularly concerned about any systematic bias in the flux-density recovery. As explained in Appendix A.2.3, we already found it necessary to force BeeP's likelihood to model beam shaped sources with a source extension EXT $\approx 1\farcm72$ (about 1 pixel). As a result, the flux-density estimation bias was much reduced but not completely eliminated. Figure B.9 shows the comparison of $\ln(S_{\rm out}/S_{\rm in})$ based on the analysis of the FFP8 maps by BeeP. The figure shows a bias in this quantity and a large dispersion, particularly at the low NPSNR regime. The median of $\ln(S_{\rm out}/S_{\rm in})$ is 0.104[55], which implies an approximately +11.0% bias in the recovered reference flux density, both for the flux-density estimates based on the MBB model and those based on the "Free" model.

The FFP8 simulations allow us to assess the impact that a realistic beam has on the positional accuracy, as was done for the injection simulations using all four channels and circularly symmetric sources (see Appendix B.2). Figure B.10 (left panel) shows that the positional accuracy of the recovery exhibits similar traits as in the case of the injection simulations. We can therefore apply the same procedure as in Appendix B.2 to correct POSERR, and we find a very similar threshold level:

$$\sigma_0 = 4\farcs43^{+0\farcs17}_{-0\farcs18}. \tag{B.7}$$

The position deviations normalized by the corrected position error bar now follow a unitary Rayleigh distribution (Fig. B.10, right panel). The median POSERR for the full FFP8 catalogue is around $13\farcs2$ (13% of a pixel), which is larger than that found in the injection simulations ($7\farcs8$, Appendix B.2). This difference must be partly due to the more realistic beam simulations included in FFP8. However, other factors are likely to play a role as well, e.g., the FFP8 catalogue has a smaller median NPSNR (9.65) when compared with the injection simulations (15.1); and the absence of the 3000 GHz channel may have an impact as well.

As in the analysis of Appendix B.2, we do not apply this correction to the BeeP output. The correction should be applied only when well-behaved statistical characterization of a sample containing high-NPNSR sources is required. In this case, we recommend to use the slightly more conservative correction value shown in Eq. (B.7).

Appendix C: The no-source simulations: frequentist versus Bayesian approaches

In Appendix A we quite closely followed a frequentist framework:

- we define the "null hypothesis" to be that no source is present (only background);
- we define a data-based statistic SRCSIG and its cumulative distribution *assuming* that the null hypothesis is true;
- we reject the null hypothesis for extreme values of SRCSIG, using a single-tailed test.

Rejecting the null hypothesis means that the data do not support, at a certain level ("tail probability"), the background-only hypothesis. Therefore one chooses the alternative hypothesis, namely a source is present. Figure 4 shows the level at which

[55] To avoid any possible distortion resulting from Eddington-type bias, we restrict the data set to high NPSNR (>40) sources only (see green points in Fig. B.9).

we reject the null hypothesis as a function of SRCSIG. We call this "contamination" because at that threshold of SRCSIG, we still expect the null hypothesis (no real source) to be true a certain fraction of the time. That fraction is given as a percentage of the total using the y-axis of Fig. 4. This is a commonly employed method of measuring the contamination of a catalogue.

When alternatively using the Bayesian framework, instead of dealing with each source individually, we prefer to address the broader question of "catalogue contamination". Catalogue contamination can be defined as the percentage of false sources in the sub-catalogue defined by a given threshold of the selection statistic:

$$\Pr(\mathrm{F}|\mathrm{C};\theta), \tag{C.1}$$

where F means a "false" source, C means "it is part of the catalogue", and θ is the threshold. Using conditional probability rules and Bayes theorem, the definition of Eq. (C.1) can be expressed as

$$\Pr(\mathrm{F}|\mathrm{C}) \equiv \Gamma = \left(1 + \frac{\Pr(\mathrm{T})\Pr(\mathrm{C}|\mathrm{T})}{\Pr(\mathrm{F})\Pr(\mathrm{C}|\mathrm{F})}\right), \tag{C.2}$$

where T means a *true* source and Pr(T) and Pr(F) are the prior probabilities (before running BeeP) of a source being real or spurious. Pr(C|F) is what we compute from the "no-source" simulation and Pr(C|T) is the completeness. We have dropped the threshold from the expression to make it more readable, although all the factors are dependent on it. The quantity Γ is a proper probability $\in [0, 1]$. Let us define

$$\alpha \equiv \frac{\Pr(\mathrm{T})\Pr(\mathrm{C}|\mathrm{T})}{\Pr(\mathrm{F})}. \tag{C.3}$$

Then Γ reads,

$$\Gamma = \frac{\Pr(\mathrm{C}|\mathrm{F})}{\Pr(\mathrm{C}|\mathrm{F}) + \alpha}. \tag{C.4}$$

If $\alpha \gtrsim 1$ then $\Gamma \lesssim \Pr(\mathrm{C}|\mathrm{F})$ and the no-source simulations give a good estimate of an upper bound on the expected catalogue contamination. It is interesting to note that the catalogue completeness is also present when computing the catalogue contamination using a Bayesian approach.

A useful catalogue must always have the following properties:

- completeness $\equiv \Pr(\mathrm{C}|\mathrm{T}) \approx 1$;
- contamination $\equiv \Pr(\mathrm{C}|\mathrm{F}) \approx 0$.

If $\Pr(\mathrm{T})/\Pr(\mathrm{F}) \approx 1$, i.e., no prior bias, then $\Gamma \approx \Pr(\mathrm{C}|\mathrm{F})$, as in a frequentist result. In the extreme case of a SRCSIG threshold of zero, then $\Pr(\mathrm{C}|\mathrm{T}) = \Pr(\mathrm{C}|\mathrm{F}) = 1$ (i.e., we accept everything) and $\Gamma = 1/2$, the value one would expect if $\Pr(\mathrm{T})/\Pr(\mathrm{F}) \approx 1$.

In Planck Collaboration XXVI (2016) (bottom right panel of Fig. 7) there are no reliability values provided for the PCCS2E. However, assuming that the PCCS2E reliability is as low as 70% at 1 Jy and $\Pr(\mathrm{C}|\mathrm{T}) \approx 0.4$, then $\alpha \approx 0.93$ and Γ (Eq. (C.4)) is $\approx \Pr(\mathrm{C}|\mathrm{F})$, just like our prediction for BeeP's catalogue contamination. When completeness is very low ($\Pr(\mathrm{C}|\mathrm{T}) \approx 0$) or $\Pr(\mathrm{T}) \ll \Pr(\mathrm{F})$, then false objects are dominant and $\alpha \approx 0$ implies $\Gamma \approx 1$. In this case, even with a good rejection of false detections, the catalogue contamination can reach very high values[56]. We cannot completely rule out this scenario at very low Galactic latitudes ($|b| < 1°$) close to the Galactic centre. In this region, the properties of PCCS2+2E are not well defined and false detections could dominate.

Appendix D: Source examples

In this appendix we show a few representative examples of SEDs resulting from the analysis of BeeP. Specifically we show:

- three archetypal nearby galaxies, Arp 220, M 100, and NGC 895 (Clements et al. 2014);
- one source (J091828.6+514223) from the *Planck* list of high-redshift candidates (Planck Collaboration Int. XXXIX 2016), also detected in the *Herschel* Lensing Survey overview (Egami et al. 2010), which is a strongly lensed galaxy at $z = 5.2$ (Combes et al. 2012);
- one source from the GEMS catalogue (*Planck* dusty Gravitationally-enhanced submillimetre sources, Cañameras et al. 2015), PLCK G138.6+62.0;
- two sources with non-thermal SEDs that cannot be fit to an MBB spectrum, namely M1 (the Crab Nebula, a supernova remnant) and 3C 273 (a blazar);
- the brightest source in our ATLAS comparison field (HATLAS J144011.1-001719);
- one of the coldest Galactic clumps extracted from the PGCC (Planck Collaboration XXVIII 2016), IRDC MSXDC G033.69−00.01;
- Orion A IRC 2, an archetypal infrared source in the Orion A molecular cloud.

[56] For an extreme (but realistic) example, see p. 1132 of Riley et al. (2006).

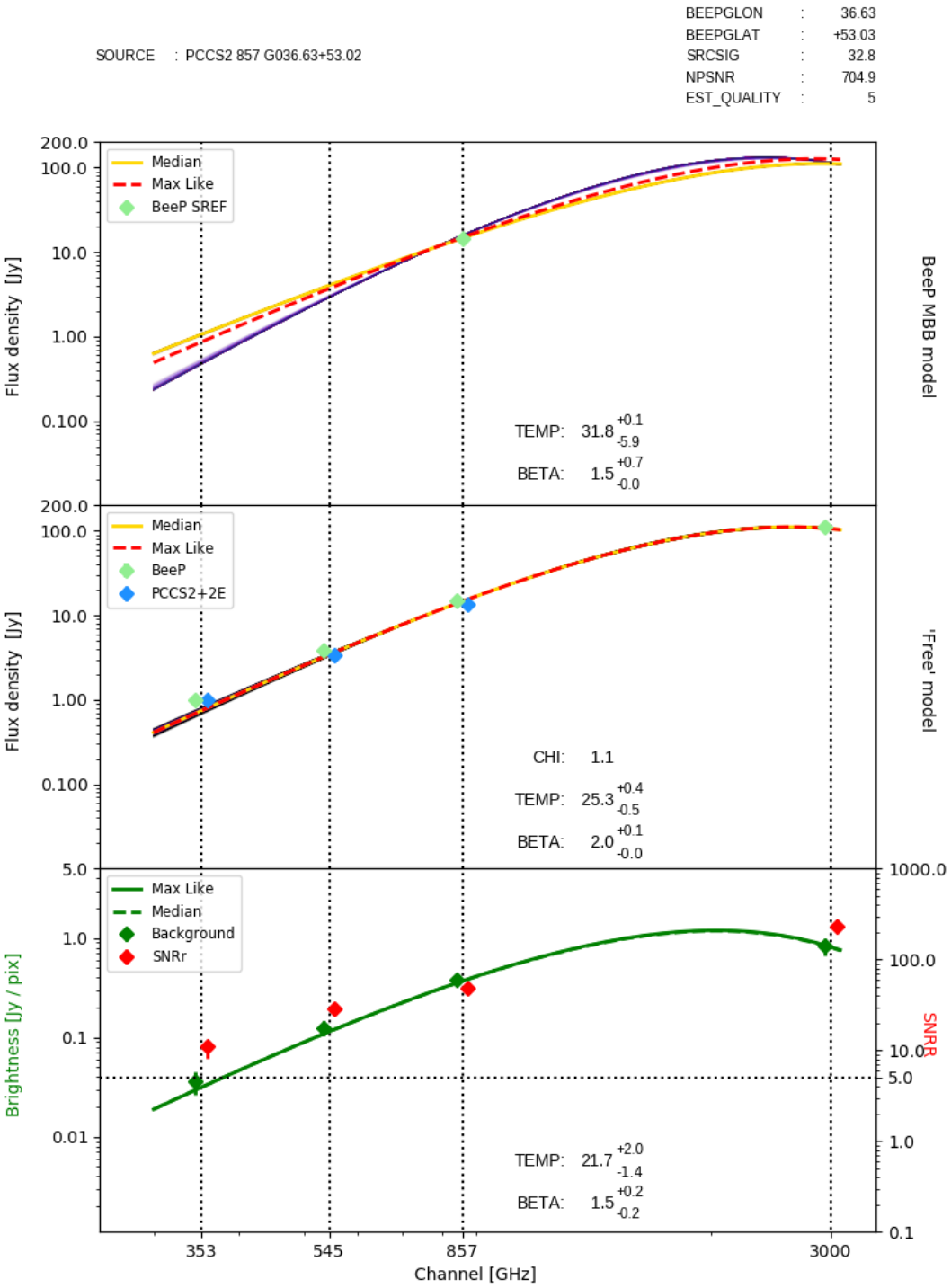

Fig. D.1. Results of BeeP analysis for Arp 220. See the caption of Fig. 7 for a full description of the contents of this figure. This case is a very clean example of a well-determined model for source and background.

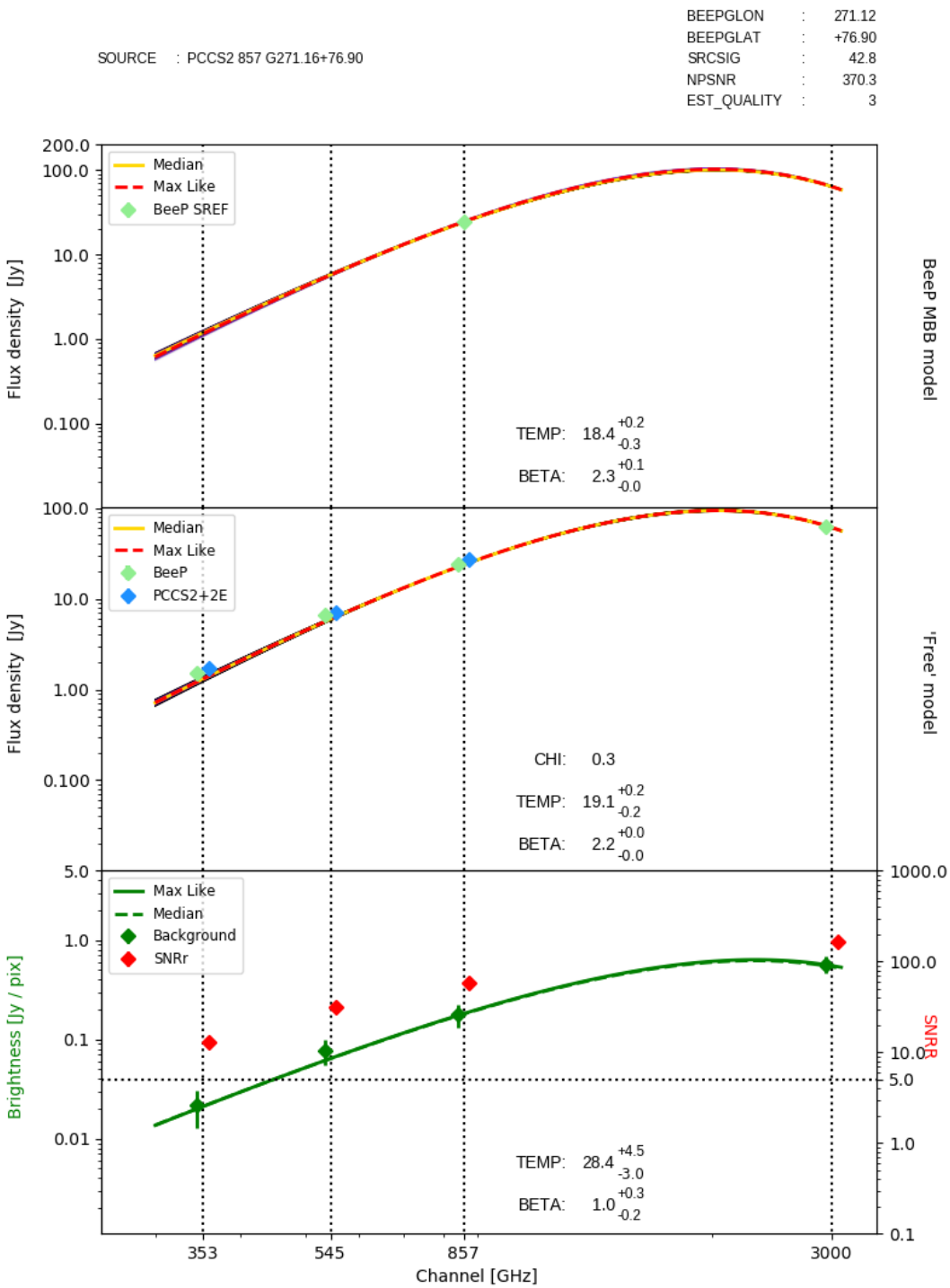

Fig. D.2. Results of BeeP analysis for M 100. See the caption of Fig. 7 for a full explanation. This case is interesting because BeeP has reduced EST_QUALITY due to the extremely low uncertainties in both temperature and spectral index (Sect. 5.2), in spite of the fact that the SEDs fit the data very well. However the χ^2 value of the Free-model fit (*middle panel*) is not far from the expected unity-per-degree of freedom level, and so this is one of those exceptional cases where the very low uncertainties reflect a very good fit, rather than the fact that the sampler has not been able to explore the parameter space.

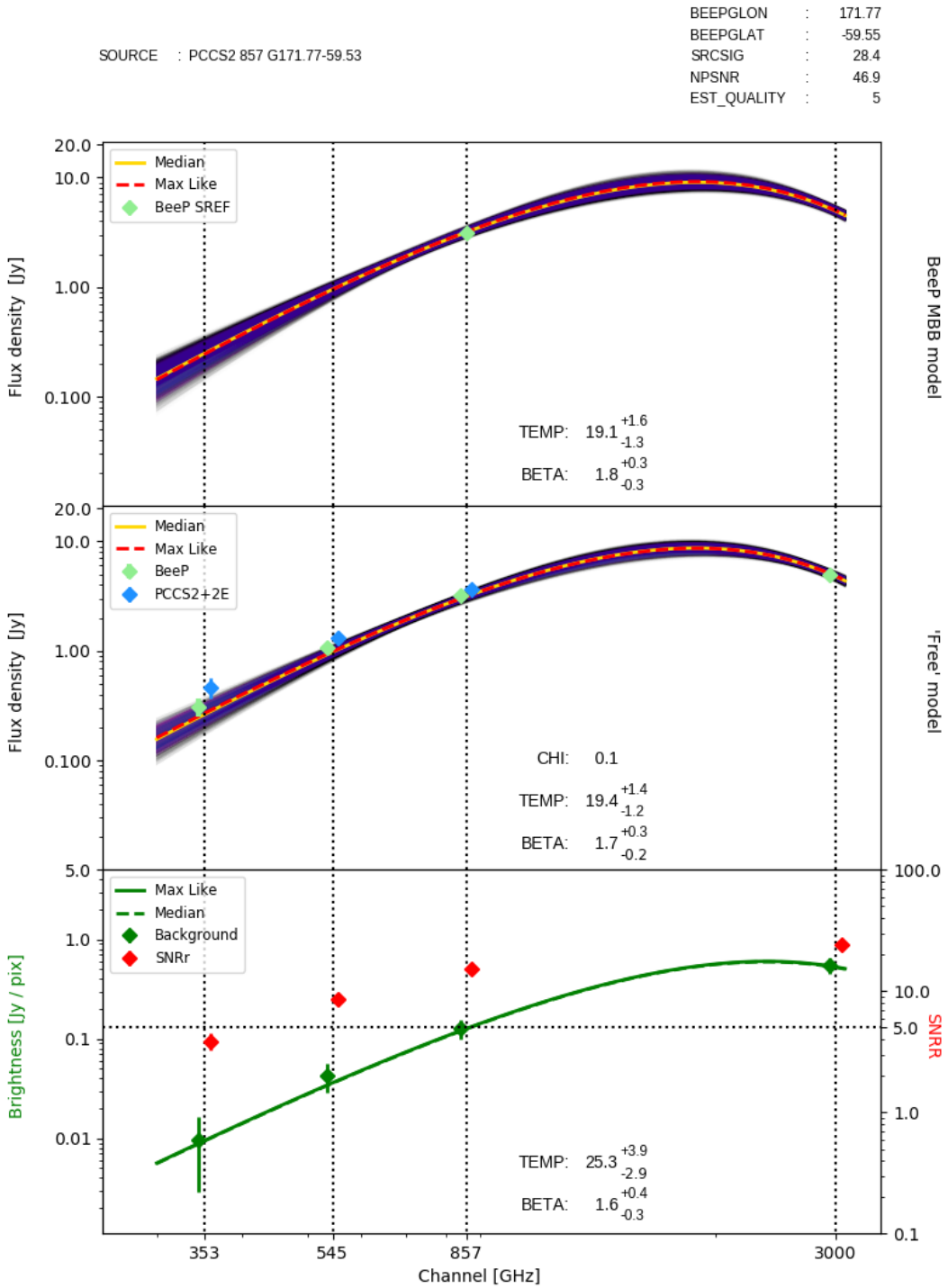

Fig. D.3. Results of BeeP analysis for NGC 895. See the caption of Fig. 7 for a full explanation. This case is also a very clean example of a well-determined model for source and background.

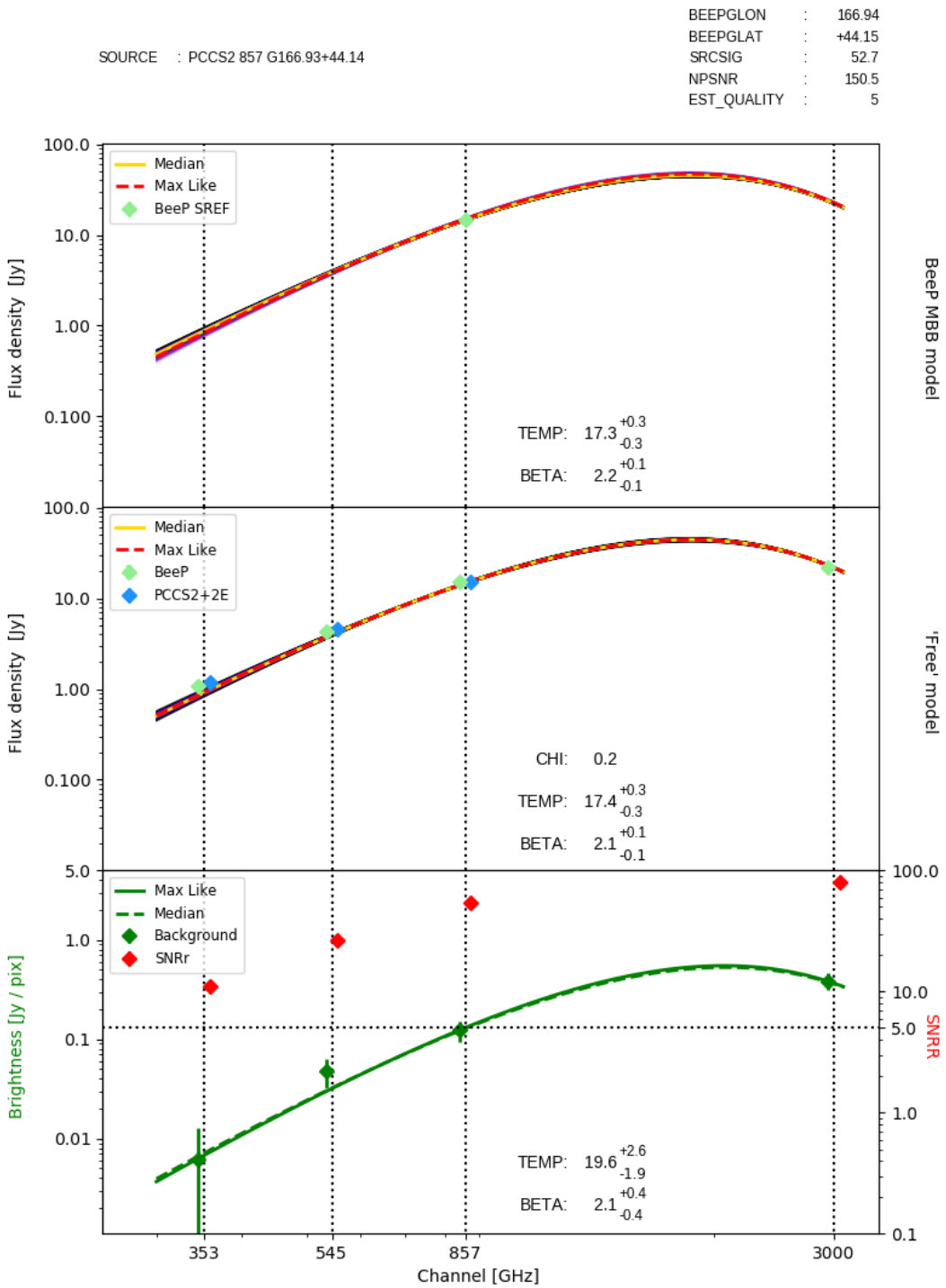

Fig. D.4. Results of BeeP analysis for J091828.6+514223. See the caption of Fig. 7 for a full description. This is a strongly lensed galaxy at $z = 5.2$ and appears as a relatively cold dusty source.

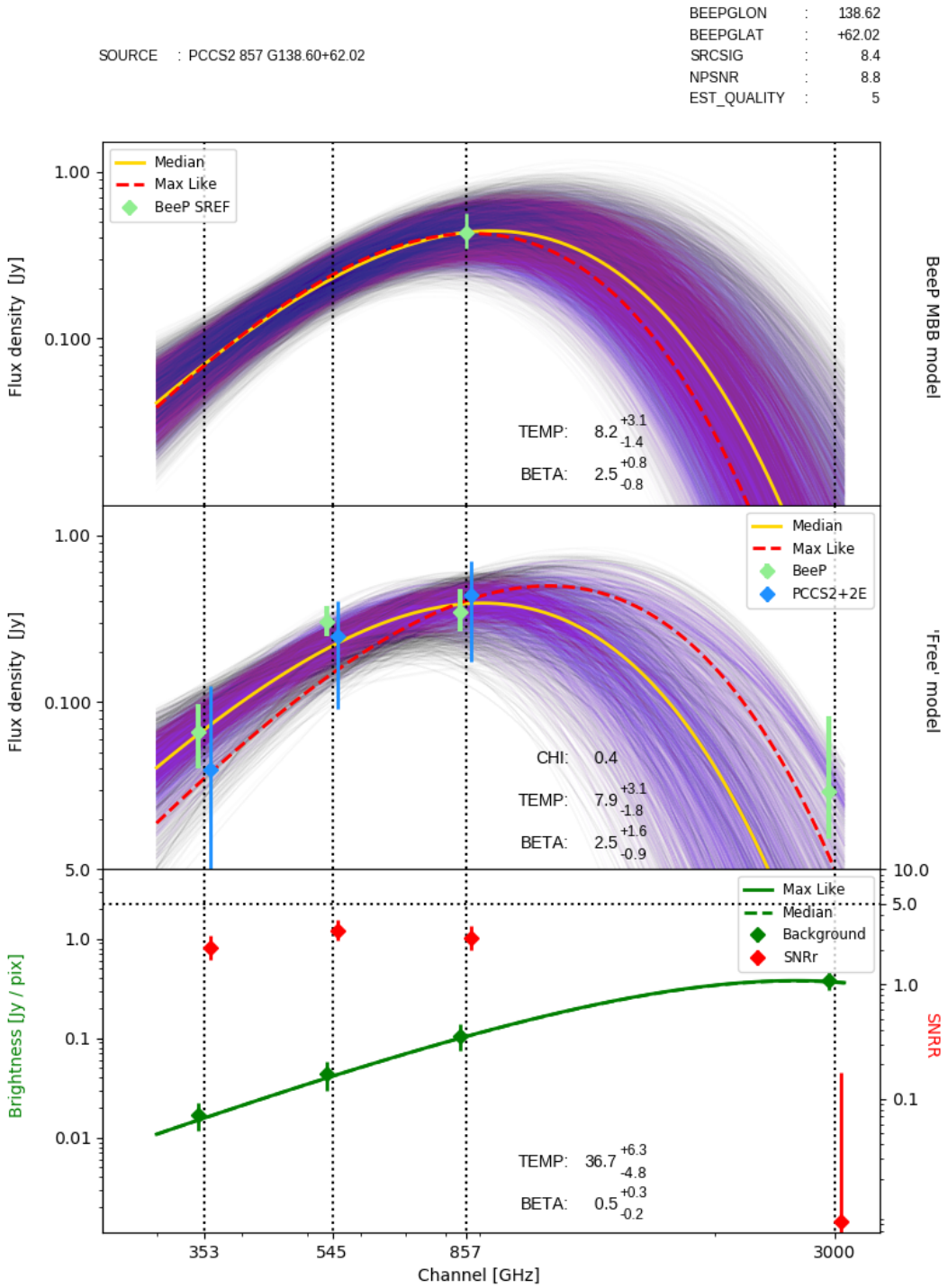

Fig. D.5. Results of BeeP analysis for GEMS PLCK G138.6+62.0. See the caption of Fig. 7 for a full description. This is a source with fairly low S/N ratio with respect to the background, but BeeP is able to find a good model for it. The Free model flux densities recovered by BeeP have much lower uncertainties than those found in the PCCS2+2E catalogue.

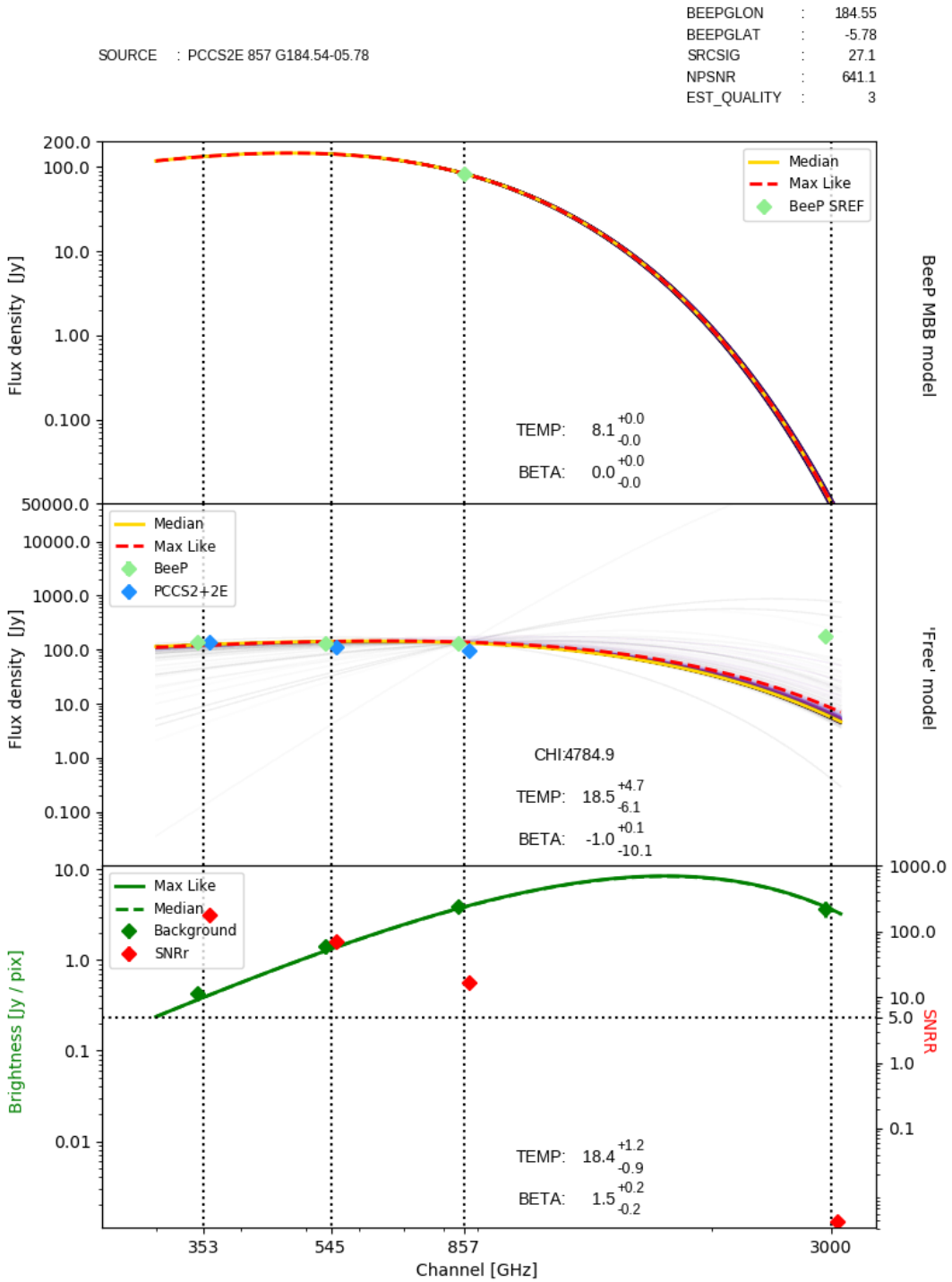

Fig. D.6. Results of BeeP analysis for M 1 (Crab Nebula). See the caption of Fig. 7 for a full description. This is a non-thermal source and BeeP has reduced EST_QUALITY accordingly; the full likelihood is not able to find a reasonable value for the SED parameters, in particular the spectral index. The Free model fit does find parameters, since it is less constrained, but the high χ^2 value indicates a very poor fit.

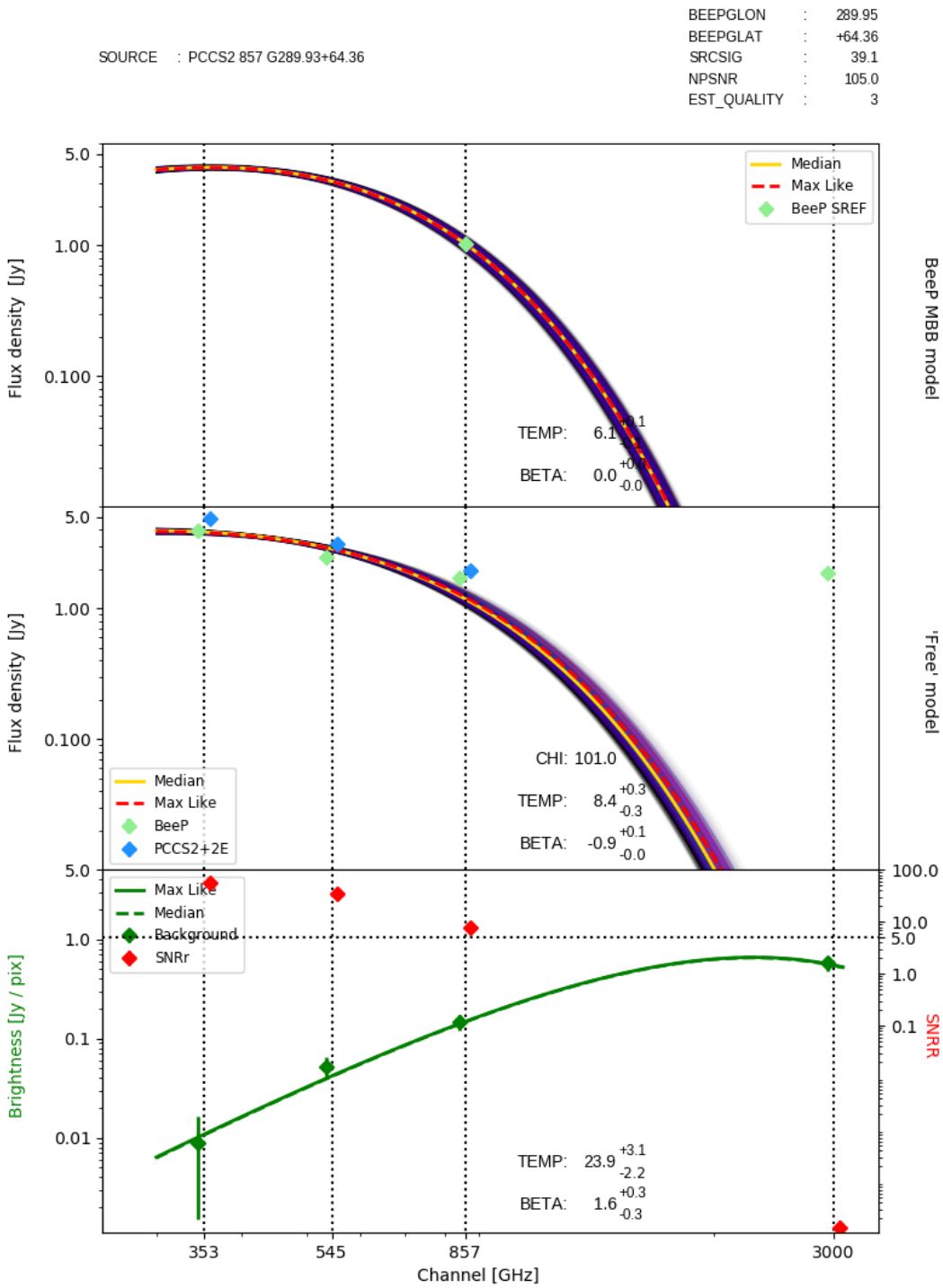

Fig. D.7. Results of BeeP analysis for 3C 273. See the caption of Fig. 7 for a full description. As in the previous figure, this is a non-thermal source, and BeeP also obtains very poor results (though not as extreme as in the previous case). The flux density of the source in the IRIS map is highly anomalous, but also has very low S/N ratio with respect to the well-determined background.

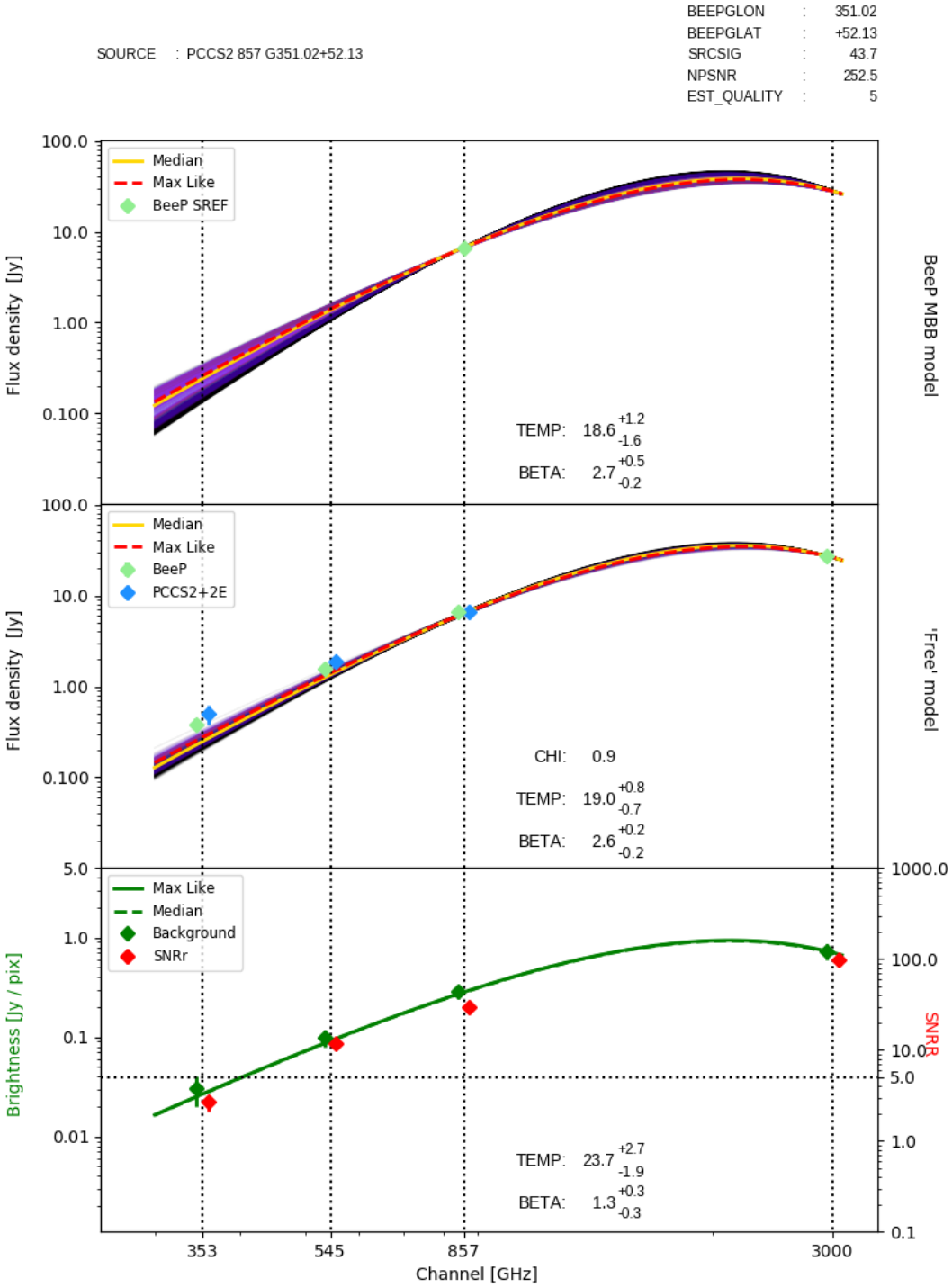

Fig. D.8. Results of BeeP analysis for the brightest source in our ATLAS comparison field (HATLAS J144011.1-001719). See the caption of Fig. 7 for a full description. Overall this is a clean case of a cold dusty source on a fairly warm background, and BeeP obtains good results.

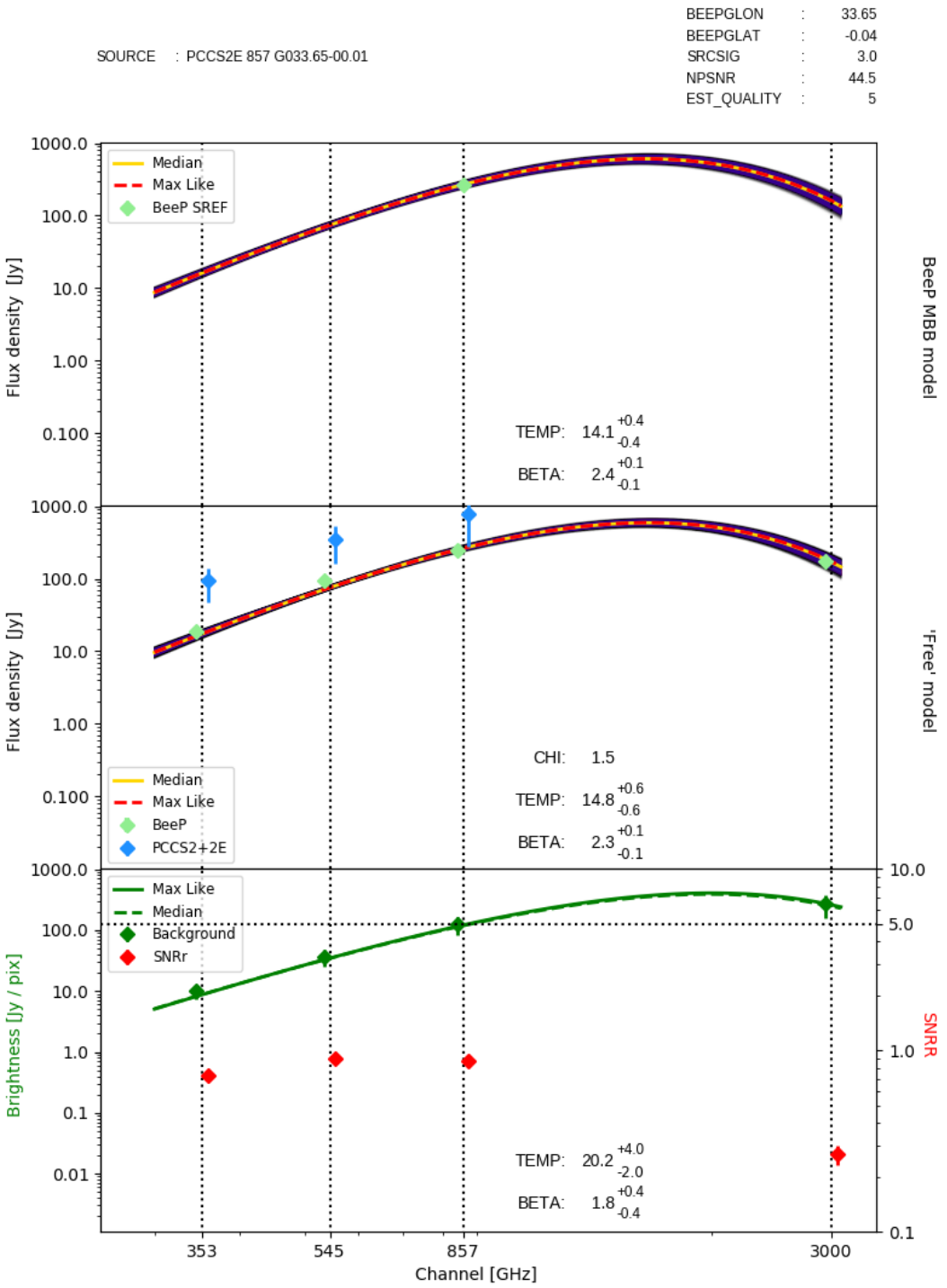

Fig. D.9. Results of BeeP analysis for cold clump IRDC MSXDC G033.69−00.01. See the caption of Fig. 7 for a full description. We note the low temperature of this source (about 14 K). BeeP obtains a good fit for this cold source on a warm background, but we see that the recovered flux densities (*middle panel*) are well below those obtained by PCCS2+2E. Examination of the source maps shows that it is surrounded by bright complex structure, which has confused the aperture photometry used by PCCS2+2E; indeed other flux-density algorithms (e.g., DETFLUX in PCCS2) obtain values closer to those of BeeP.

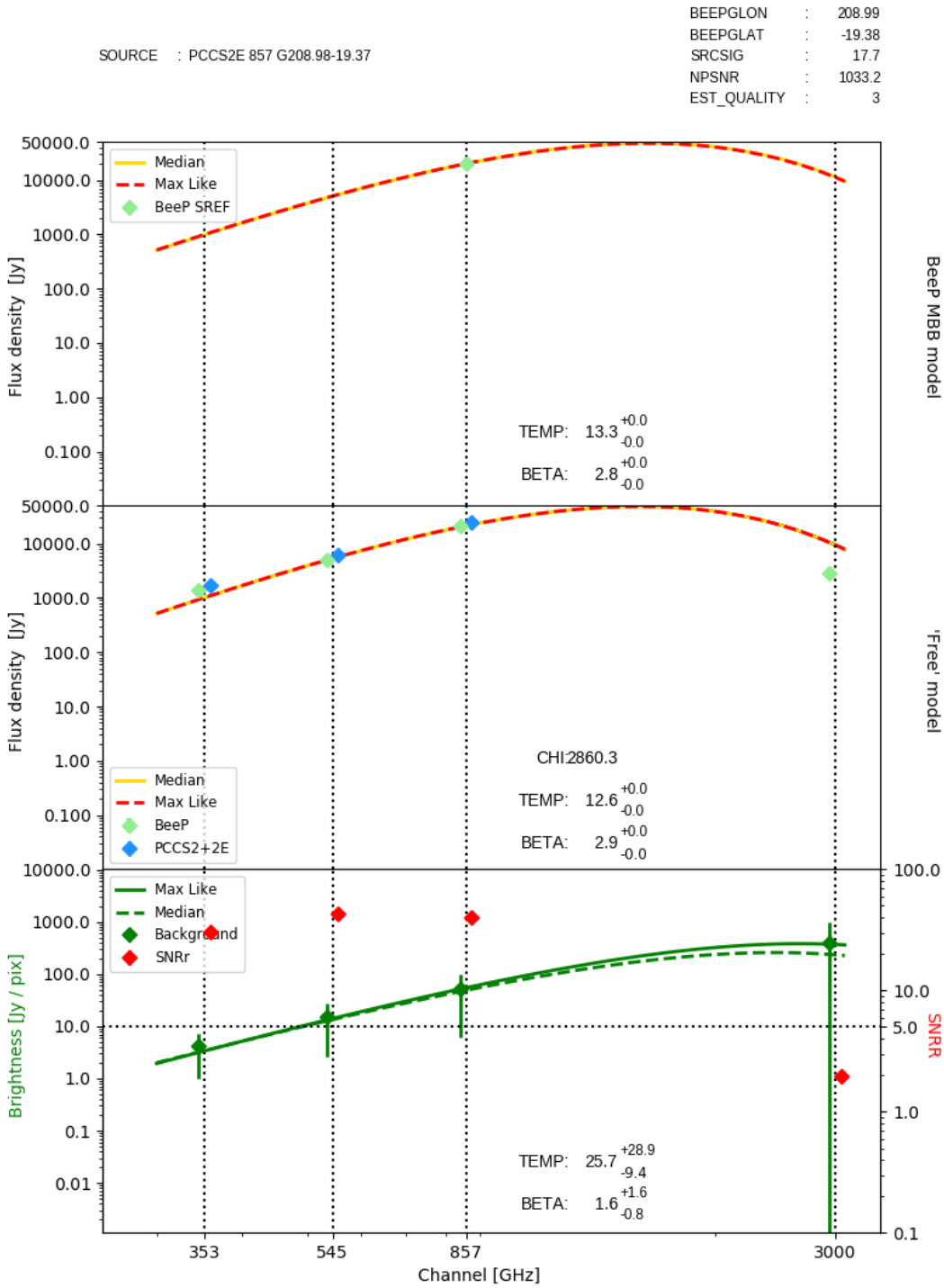

Fig. D.10. Results of BeeP analysis for Orion IRC 2. See the caption of Fig. 7 for a full description. This case is similar to that of Fig. D.2, where BeeP has reduced EST_QUALITY due to the very low parameter uncertainties. However, in this case the χ^2 of the Free model fit is very high, and this is clearly due to the fact that very tight constraints coming from the *Planck* data do not allow a satisfactorily fit to the low flux density in the IRIS map.