Point Source Detection with Fully-Convolutional Networks: Performance in Realistic Simulations

Bonavera L.1, 2, Suárez Gomez S. L.2, 3, González-Nuevo J.1, 2, Santos J.D.1, 2, Sanchez M.L.1, 2, Muñiz R.5, de Cos F.J.2, 5

1Departamento de Física, Universidad de Oviedo, C. Federico García Lorca 18, 33007 Oviedo, Spain
2Instituto Universitario de Ciencias y Tecnologías Espaciales de Asturias (ICTEA), C. Independencia 13, 33004 Oviedo, Spain
3Departamento de Matemáticas, Universidad de Oviedo, C. Federico García Lorca 18, 33007 Oviedo, Spain
4Departamento de Informática, Universidad de Oviedo, Edificio Departamental 1. Campus de Viesques s/n, 33204 Gijón, Spain
5Escuela de Ingeniería de Minas, Energía y Materiales, Independencia 13, 33004 Oviedo, Spain

November 28, 2019

ABSTRACT

Context. Point Sources are one of the main contaminants to the recovery of Cosmic Microwave Background signal at small scales, and their careful detection will be important for the next generation of Cosmic Microwave Background experiments like CORE.

Aims. We want to develop a method based on Fully Convolutional Networks to detect sources in realistic simulations and compare its performance against one of the most used point source detection method in this context, the Mexican Hat wavelet 2 (MHW2). The frequencies for our analysis are the 143, 217 and 353 GHz Planck channels.

Methods. We produce realistic simulations of Point Sources at each frequency taking into account potential contaminating signals as the Cosmic Microwave Background, the Cosmic Infrared Background, the Galactic thermal emission and the instrumental noise. We first produce a set of training simulations a 217 GHz to train the network. Then we apply both the neural network and the wavelet to recover the point sources in the validating simulations at all the frequencies, comparing the results by estimating the reliability, completeness and flux density estimation accuracy.

Results. In the extra-galactic region with a 30◦ galactic cut, the neural network successfully recover point sources with 90% of completeness corresponding to 300, 139 and 227 mJy for 143, 217 and 353 GHz respectively. On the same validation simulations, the wavelet with a 3σ flux density detection limit, recover point sources till 224, 124 and 154 mJy at the 90% completeness. To reduce the amount of spurious sources, we also apply a safer 4σ flux density detection limit increasing the 90% of completeness levels: 298, 173 and 227 mJy. In all cases the neural network produce a much lower number of spurious sources with respect the MHW2. As expected, the results on spurious sources for both techniques worsen when increasing the frequency or reducing the galactic cut to 10◦.

Conclusions. Our results suggests that the neural networks are a very promising approach to detect point sources using data from Cosmic Microwave Background experiments, providing overall better results with respect to the more usual filtering approaches.

Key words. Techniques: image processing – cosmic background radiation – Submillimeter: galaxies

1. Introduction

The importance of compact sources (galaxy clusters and extra-galactic sources) for ground- and space-based Cosmic Microwave Background (CMB) experiments has been clear since the conception of the WMAP (Bennett et al. 2003) and Planck (Planck Collaboration et al. 2013b) missions. Point sources (hereafter PS) in the microwave regime are mainly blazars (i.e. AGNs with the relativistic jets aligned along the line of sight) and dusty galaxies. At such frequencies, PS are one of the contaminants to the recovery of the CMB anisotropies signal whose effect is more important at small angular scales. For this reason PS are even more important for the next generation of CMB experiments with higher resolution than Planck, such as the Cosmic Origins Explorer (CORE, Delabrouille et al. 2018), the Probe of Inflation and Cosmic Origins (PICO, Hanany et al. 2019) or the Lite satellite for the studies of B-mode polarization and Inflation from cosmic background Radiation Detection (LiteBird, Matsumura et al. 2014). Generally, they are planned to keep the PS contamination low, but precise CMB measurements will still be affected by PS. For this reason, it is quite important to develop highly performing methods for PS detection.

The standard single-frequency approach for PS detection in the CMB and far IR frequencies rely on the Mexican Hat Wavelet (MHW; Vielva et al. 2003, González-Nuevo et al. 2006) or on the matched filter techniques (Tegmark & de Oliveira-Costa 1998, Barreiro et al. 2003, López-Caniego et al. 2006, Herranz et al. 2002). Matched filter is theoretically the optimal filter when the PS shape is known providing the maximum signal-to-noise amplification. However, as concluded by López-Caniego et al. (2006), the second member of the MHW family (MHW2; González-Nuevo et al. 2006) provides a similar performance as the Matched Filter one, but it is easier to implement and more robust. Such wavelet has been successfully applied to Planck realistic simulation (González-Nuevo et al. 2006, López-Caniego et al. 2006, Leach et al. 2008) as well as to WMAP (López-Caniego et al. 2007, Massardi et al. 2009) and Planck real data: the Early Release Compact Source Catalogue (ERCCSC, Planck Collaboration et al. 2011b), the Planck Catalogue of Compact Sources (PCCS, Planck Collaboration et al. 2014b) and the Second Planck Catalogue of Compact Source (PCCS2, Planck Collaboration et al. 2016b). This is why we decide to compare our results against such method.
Although there was always a tight relationships between Machine Learning techniques and astrophysics/cosmology, in the recent years the particular usage of neural networks has become a mainstream technique to derive new results. Artificial Neural Networks are Artificial Intelligence (AI) techniques involving numerical mathematical models which can be trained to represent complex physical systems by supervised or unsupervised learning (Gómez et al. 2019; Suárez Gómez et al. 2017). This characteristics are perfect to provide further results in Cosmology. Some example of recent interesting applications of Artificial Neural Network in cosmology are the identification of galaxy mergers (Pearson et al. 2019) and strongly gravitational lenses (Petrillo et al. 2017) in astronomical images, a better estimation of cosmological constrains from weak lensing maps (Fluri et al. 2019) and high fidelity generation of weak lensing convergence maps (Mustafa et al. 2019) and cosmological structure formation simulations under different assumptions (Mathuriya et al. 2018; He et al. 2019; Perraudin et al. 2019; Guisarma et al. 2019).

Some AI approaches, such as Multi-Layer Perceptron (MLP) (Juez et al. 2012), or Convolutional Neural Networks (CNNs) (Suárez Gómez et al. 2019; Krizhevsky et al. 2012), have been successfully applied to image processing (and related fields) for modelling and forecasting (Graves et al. 2013; Giusti et al. 2013). In this work, we propose the use of Fully-Convolutional Networks (FCN) (Long et al. 2015; Dai et al. 2016) as a very promising tool for PS detection. They are usually applied in image recognition and make use of different layers in order to get various image features (e.g., shapes, smoothness and borders). Important features of the input images are generally obtained by pairing convolution and merging layers. After that, layers are applied to get the output (image or numerical). In this work we present an application of FCN to the detection of PS in realistic simulations, the Point Source Image Detection Network (PoSeIDoN) that can be summarised in the search of spheroids in a noisy background (i.e. the rest of the components in the microwave sky).

The outline of the paper is the following. Section 2 describes how the simulated maps are generated and Section 3 reviews our methodology. The results are presented in Sections 4 and our conclusion are in Section 5.

2. Simulations

In this work, we make use of realistic simulated maps of the microwave sky. The simulations correspond to sky patches at 143, 217 and 353 GHz, the central channels of the Planck mission, with $\text{pixsize} = 90 \text{ arcsec}$ (a round number close to the 1.72 arcmin used in the Planck maps) and $\text{npix} = 2048$ in the HEALPIX all-sky pixelization schema (Górski et al. 2005). For memory and speediness reasons, we use patches of $128 \times 128$ pixels, a trade of between density of bright sources per patch and size. We tested anyway that using bigger patches ($256 \times 256$) does not alter our statistical results or our conclusions.

First, a catalogue of radio PS is simulated at each frequency independently by following the model by Tucci et al. (2011). The flux density limit is $1 \text{ mJy}$ at all the frequencies. From the simulated catalogue we then create the simulated PS map and convolve it with the FWHM of the instrument (7.22, 4.90 and 4.92 arcmin at 143, 217 and 353 GHz respectively; Planck Collaboration et al. 2018). In order for the simulations to be realistic at these frequencies, we need to take into account fluctuations due to high redshift infrared PS (massive proto-spheroidal galaxies in the process of forming most of their stellar mass; Granato et al. 2004; Lapi et al. 2006; 2011; Cai et al. 2013) too faint to be detected one by one. Such contamination (Blain et al. 1999; Lagache et al. 2003; Dole et al. 2004) is dominant at few arcmin resolution and it is called the Cosmic Infrared Background (CIB; Puget et al. 1996; Hauser & Dwek 2001; Dole et al. 2006). We use the software CORRsky (González-Nuevo et al. 2005) to simulate a sample of galaxies with a particular clustering properties, described by their angular power spectrum $p(k)$. We adopt the power spectrum and the source number counts (different at each frequency) given by the Lapi et al. (2011) and Cai et al. (2013) models.

As the main idea of this work is to compare detection methodologies, we do not simulate the late-type infrared galaxies (Tofolatti et al. 1998; Planck Collaboration et al. 2011a, 2013, 2014b), that dominate the bright part of the source number counts above 217 GHz. Radio and late-type galaxies are only distinguishable for their different spectral emission, not for their shape and/or size. Compared with the Planck beam they are both point like sources. Their introduction will have supposed simply a higher density of brighter PS per patch in the highest simulated frequency without appreciably modifying the statistical properties of the background.

On larger angular scales, we must include in our simulation the contamination due to diffuse emission by our Galaxy and the CMB. Such contaminants are introduced in our simulated maps by randomly select patches in Planck 143, 217 and 353 GHz official CMB maps (from the last release described in Planck Collaboration et al. 2018b). The CMB maps are the one by the SEVEM method (Leach et al. 2008; Fernández-Cobos et al. 2012), that are provided at all Planck frequencies. For the Galaxy emission we use the Planck FFP10 simulation available for all Planck channels. The Planck maps are at $\text{npix}=2048$, that corresponds to a pixel size of 1.72 arcmin, and the selected sky patches are projected into flat patches with pixel size of 1.5 arcmin using the gnomview function provided with the HEALPix framework (Górski et al. 2005).

Finally, we add the instrumental noise to the simulations. The noise maps are produced by simulating white noise accordingly to the Planck values: 0.55, 0.78 and 2.56 $\mu \text{K}_\text{CMB}$ deg. respectively (Noise rms computed after smoothing to 1º; Planck Collaboration et al. 2018b).

In this work, we study the performance of two detection methods, PoSeIDoN and the MHW2, especially focusing on their dependence with increasing intensity of Galactic emission by applying two different homogeneous galactic cuts (at 10º and 30º galactic latitudes). Moreover, such intensity increase also arises with higher frequencies due to Galactic emission spectral behaviour (Planck Collaboration et al. 2011c, 2014a, 2016a, 2018c).

Examples of random simulated patches are shown in the first two columns of Fig. 1 for 143, 217 and 353 GHz (top, middle and bottom panels, respectively). The first column is the total input simulated map, including CMB, Galactic emission, CIB, PS and instrumental noise, whereas the second column is the input PS only map.

3. Methodology

3.1. PoSeIDoN

Neurons, sorted in layers, are the basic computing elements of an artificial network model. Their responses are modelled by

\[ \text{ available at http://pla.esac.esa.int/pla/} \]
Fig. 1. From left to right, sample patch comparison among the total and PS input validation maps and the MHW2 and PoSeIDoN PS outputs, for 143, 217 and 353 GHz (top, middle and bottom panel respectively). The number, position and flux density of the PS are different at each frequency.

Fig. 2. Details of the FCN used for PS detection in PoSeIDoN. The network has a block of 8 convolutional layers, where the main characteristics are extracted, resulting in 512 feature maps, connected with a deconvolutional block of 8 deconvolutional layers. Fine-grained features are added from each convolution to the correspondent deconvolution.

weights that represents the influence of the neuron response on the neurons of the subsequent layer. In particular, for some models (such as CNNs) the weights correspond with kernel values [LeCun et al. 2015]. The response is finally given after the process is completed along each computation units.

In supervised learning, the implementation of the training procedure is performed via estimation of a loss function, usually a Mean Square Error (MSE) function, computed over the data from a training set (i.e. the network responses to certain inputs compared with their corresponding labels). Back-propagation algorithms are then employed to correct weights and kernel values and thus minimise the loss function with methods as the Stochastic Gradient Descent (SGD) [Rumelhart et al. 1988; Chauvin & Rumelhart 2013].

FCNs allow us to perform dense predictions over the data used as input [Long et al. 2015; Dai et al. 2016]. In this case, the most relevant characteristics are first extracted using a convolutional block where each convolutional layer allows the extraction
of several feature maps from the image obtained using kernels, frequently modulated by an activation function and processed by a down-sampling in terms of pooling. In addition to the typical convolutional process, an FCN has a second block where deconvolutions are performed, allowing the recovery of a dense response, also by means of layers with the correspondent kernels. Moreover, during the deconvolution process, information on the convolutional segment is included through the addition of fine-grained features in specific steps.

In this work, the FCN parameters and hyperparameters are selected through a grid search. The selected topology is detailed as follows (see Figure 2).

- Convolutional block: the network has six convolutional layers, with 8, 2, 4, 2, 2 and 2 kernels respectively. Their correspondent kernel sizes are of 9, 9, 7, 7, 5 and 3 values of side. The activation function is leaky ReLU \( \text{ReLU}(\text{Nair & Hinton} \ 2010) \) in all the layers. Strides are of pixels both horizontally and vertically and padding has been added.

- Deconvolutional block: the feature maps obtained after the convolutions are connected to a block of six deconvolutional layers. These layers have 2, 2, 2, 4, 2 and 8 kernels respectively. Their correspondent kernel sizes are of 3, 3, 5, 7, 9 and 9 values of side. The activation function is leaky ReLU in all the layers. Strides are of pixels both horizontally and vertically and padding has been added. Moreover, feature maps resulted from the five last convolutions are added, as fine-grained features, to the results of the five first deconvolutions.

The training procedure is performed using an MSE loss function, with a training set of 50000 samples and a validation set of 5000 samples. The test sets for performance assessment consists of 5000 samples too.

We produce 50000 simulations at 217 GHz to train the network. For each simulation we randomly chose a position of the available sky with the selected cut in latitude (10° or 30°) for both the CMB and the galactic emission. Moreover, the positions and fluxes for the input PS are also different in each realisation. At this stage, for each patch, two images are provided to PoSeIDoN: the total image (the simulated patch including all the components, the "Input Total" column in Fig. 1) and the PS image (the image containing only the input PS that should be detected; the "Input PS" column in Fig. 1). Mind that just for the training procedure, the sources flux density in the simulated catalogue are amplified by a "training factor" of 10, before being added to the other components. The reason is simply to increase the density of possible bright PS inside the patch without modifying the source number count shape, i.e. without altering the statistical properties of the PS sample, just their normalisation.

Please note that PoSeIDoN is trained just at 217 GHz and for the 30° galactic cut. Such trained FCN is then applied to all the cases studied in this work (i.e. 143, 217 and 353 GHz with a 30° Galactic cut and 217 GHz with a 10° Galactic cut). Better results are expected, although probably modest ones, if PoSeIDoN can be trained at each case individually. On this respect, the detection of PS in regions with intense Galactic emission, as the Galactic plane, is probably the most interesting case and also the one that can be improved the most by a dedicated FCN training. However, this is beyond the main scope of the current work.

On the other hand, in the validation process, the simulated sources flux densities are the realistic ones (no additional training factor is applied), that also allow us to compare our results with the Planck catalogues. The validation simulation is built using realistic PS flux densities and realistic contaminants simulated in the same way as for the training ones (although the sky positions are always randomly chosen). Each validation patch is then provided to the trained network that returns an output map of recovered PS. An example of the PoSeIDoN output patch at the studied frequencies is shown in Fig. 1 last column. Such output is then compared with the input PS only map for a performance analysis: estimation of the completeness, reliability and flux density accuracy.

### 3.2. Mexican Hat Wavelet 2

To assess PoSeIDoN performance, we also compare it against the MHW2 filter. The Mexican Hat Wavelet Family in the plane is derived by applying the Laplacian operator iteratively to the 2D Gaussian (González-Nuevo et al. 2006). Any member of the family can be written in Fourier space as:

\[
\psi_n(k) = \frac{k^2 e^{-k^2/2}}{2^n n!}
\]

The first member of the family, \( \psi_1 \), is the traditional MHW. It is one of the first wavelets applied successfully to the detection of PS in flat CMB maps (Cayón et al. 2000; Vielva et al. 2001). The MHW2 is therefore the second member of the family, \( \psi_2 \), and it was demonstrated even more suited to the task than its predecessor (González-Nuevo et al. 2006) or the theoretical optimal Matched Filter (Tegmark & de Oliveira-Costa 1998). In fact, it was successfully applied to the WMAP data (González-Nuevo et al. 2008; Massardi et al. 2009) and it became the standard filtering technique for the production of the PS catalogues for the Planck mission (Planck Collaboration et al. 2011b, 2014b, 2016b).

The wavelets coefficients, \( w_n(b, R) \), can be obtained for each member of the family as:

\[
w_n(b, R) = \int dke^{-k^2/2} f(k) \psi_n(kR),
\]

with \( b \) being the location and \( R \) the wavelet scale.

By definition, a PS adopt the beam profile or point spread function, usually approximated by a Gaussian:

\[
\tau(x) = \frac{1}{2\pi\sigma_b^2}e^{-x^2/\sigma_b^2},
\]

where \( \sigma_b \) is the instrumental Gaussian beam dispersion. Therefore, the intensity of each source can be written as:

\[
I(x) = I_0 e^{-\pi x^2/\sigma_b^2}.
\]

Then, the scale R of the wavelet can be optimised by finding the maximum signal-to-noise ratio (S/N) of the sources in the filtered patch, i.e. maximising the amplification factor \( \lambda_n = w_n\sigma_{\text{rms}}, \) with \( \sigma \) and \( \sigma_{\text{rms}} \) the deviation of the background before and after filtering, respectively. The optimal scale is determined for each patch independently and it is always near to unity.

The wavelets can be used for blind source detection (no prior information on the sources’ positions) and in non-blind mode, usually to get the estimated flux densities of PS at known positions. In our case, we apply the filter blindly to each total input validation simulation to produce the filtered image. Examples of the MHW2 output image for our frequencies are the patches shown in Fig. 1 third column.

Article number, page 4 of 9
3.3. Catalogue production and statistical comparison

Both PoSeIDoN and the MWH2 methods provide just an output image, not a list of detections. In this section we describe the catalogue production process and the statistical quantities that we use for the performance comparison.

The catalogue extraction consists simply in searching peaks, i.e. local maxima, above a certain intensity threshold, separated by at least a given minimum distance. This distance is 1.5 times the instrumental Gaussian beam dispersion or \( \sigma_b \), that can be different for each channel.

Taking into account that the MHW2 is the most used technique to detect PS in these kind of images, we use it as our reference for the comparison. In particular, we use the standard deviation, \( \sigma_{\text{MHW2}} \), of the MHW2 output map to set up the thresholds for catalogue production. In the case of the MHW2 we set a \( 3\sigma_{\text{MHW2}} \) threshold to build the catalogue of detected PS (positions, flux densities and uncertainties). To reduce the number of spurious sources we also apply the MHW2 with the flux density threshold set to \( 4\sigma_{\text{MHW2}} \), reducing the completeness, as shown in section 3.3.1. This last threshold level is more or less the one used for the Planck official PS catalogues (Planck Collaboration et al. 2011b, 2014b, 2016b).

The input catalogue is built from the input PS only image. Taking into account that this image does not have noise or any kind of background we use a lower PS threshold for the input catalogue. We choose one \( \sigma_{\text{MHW2}} \) in order to be sure to have fainter PS. In the PoSeIDoN case, the output map is an attempt to mimic the input PS image. It contains just PS candidates without any background or instrumental noise residuals. Therefore, we apply to the FCN output map the same procedure followed for the input catalogue, i.e. the one \( \sigma_{\text{MHW2}} \) threshold.

A well known issue of any filtering technique is the border effects. As can be seen in the MHW2 patches from Fig. 1 (third column), the filtering procedure produces artefacts near the patch border that can introduce spurious PS detections. In order to make a fair comparison with the MHW2 results, we exclude the detections within 5 pixels (~ 1 FWHM) from the patches’ borders on every side for all the cases (input, MHW2 and PoSeIDoN).

To describe the performance of the two techniques we focus on three statistical quantities: completeness, number of spurious detections and flux density estimation. These statistical quantities are commonly used to validate a detection technique or a produced catalogue (see e.g., López-Caniego et al. 2007; Planck Collaboration et al. 2011b, 2014b, 2016b, 2018a; Hopkins et al. 2015).

Completeness is estimated by cross-matching the detected PS against the input catalogue. It is a function of the intrinsic flux density, the detection threshold and sky location (not analysed in this work because the location dependency is common to both techniques by using the same simulations). Completeness provides information about the cumulative number of input sources that are missed at fainter flux densities: 

\[
C(S_0) = \frac{N_{\text{det}}(S_0)}{N_{\text{input}}(S_0)},
\]

with \( S_0 \) the input flux density.

All the detection techniques misidentify background fluctuations as PS at faint flux densities or around positions with strong Galactic emissions as the Galactic plane. Those wrongly detected PS, that are not in the input catalogue, are called spurious sources. Depending on the background characteristics and intensity, in many occasions it is the number of spurious sources, and not the completeness, that put a lower limit to the minimum flux density achievable with a given detection method. Therefore, the number of spurious sources is another important statistical quantity for the performance assessment of a detection technique.

Finally, the third statistical quantity is the flux density estimation. For those input PS detected, we can compare their flux densities. This comparison can provide useful information about potential flux density bias or to identify spurious sources that were detected by chance on the same positions of faint input PS.

4. Results

As mentioned above, Fig. 1 shows examples of output maps provided by the MHW2 and PoSeIDoN techniques when applied to the simulations. The third column corresponds to the MHW2, where it can be appreciated the typical granule background after filtering. Moreover, these patches clearly show the border effects produced by the filtering approach. The fourth column are instead the PoSeIDoN output maps: no background fluctuations are present here, as the FCN only provides the best guess for input PS. It must be stressed that no border effects are present in the PoSeIDoN output images.

As explained in section 3.1, we train the FCN only at 217 GHz for a Galactic mask of \( |b| > 30^\circ \) and we apply it to all the other studied cases. The performance of both techniques at this frequency are compared in the middle panel of Fig. 3 for the the 30° Galactic cut. The completeness (top sub-panel) and the percentage of spurious sources with respect to the input ones (bottom sub-panel) are shown on the left. The relative errors of the flux density estimation \( \Delta S/S_0 = (S_{\text{det}} - S_0)/S_0 \) are shown on the right.

With the MHW2 we obtain the expected results: we have general agreement with previous applications of the MHW2 and in particular with the Planck catalogues. Using a \( 3\sigma \) threshold (blue dashed line) the MHW2 provides good completeness results with a 90% completeness level at 124 mJy. However, such aggressive threshold implies a spurious PS detection problem already at ~ 400 mJy with more than 20% of the total detected sources being spurious ones. In fact, using a more conservative and traditional \( 4\sigma \) threshold (cyan dot-dashed line) the spurious problem is highly reduced at least until ~ 200 mJy. The price to be paid for this improvement is a reduction of the 90% completeness level that increases to 173 mJy.

On the other hand, PoSeIDoN (red solid line) has a similar completeness performance with an intermediate 90% completeness level at 139 mJy. The clear advantage of the FCN is in the much lower number of spurious PS: in this case, it starts to be an issue only below ~ 100 mJy (a flux density level below the 90% completeness level). The spurious PS issue is strongly related with high intensity regions of the background (mainly the Galactic emission). Therefore, the fact that we have a lower number of spurious PS implies that PoSeIDoN is most robust in distinguishing a PS from a background local maxima.

The flux density accuracy, right column in Fig. 3 (central panel for 217 GHz), provides additional information in the understanding of the better PoSeIDoN performance when dealing with spurious PS.

On the one hand, in the MHW2 case, most of the flux densities are correctly recovered within a 10% relative error. For the \( 3\sigma \) case the relative flux density error distribution shows a strong tail toward positive error, i.e. there are many sources whose flux density is overestimated by factors > 50%. Usually, those cases correspond to spurious PS, caused by strong Galactic emission, that by chance are near the position of a very faint input PS that should not be detectable. As expected, this issue almost disappeared with the more conservative \( 4\sigma \) threshold.
On the other hand, PoSeIDoN behaviour on the recovery of the flux densities side is completely the opposite. The FCN recovers correctly the flux density of the most bright sources, but tends to under-estimate the flux density of fainter sources. A way to understand this specific behaviour is to consider that the final flux density recovered by PoSeIDoN is multiplied by a ‘confidence’ factor: $S_{\text{est}} \propto p_{\text{conf}}S_0$. For bright input PS or those in low background fluctuations areas, $p_{\text{conf}} \sim 1$. On the contrary, for faint input PS or those near high background fluctuations areas, $p_{\text{conf}} < 1$.

This ‘confidence’ factor has the advantage to put the most dubious detected PS at fainter flux densities (see the steep increase of spurious sources below $\sim 100$ mJy). But it also means that the FCN recovered flux densities are not reliable. Although this is not ideal, it is not a limitation at all in the application of this novel technique: it is not unusual to firstly apply one technique for detection and then a second different one on the detected positions to estimate the flux density with better accuracy. One simple pipeline might consist in estimating the flux densities with the aperture flux, or with the MHW2 in non-blind mode, on the PS positions provided by PoSeIDoN. Another interesting possibility is to train a second neural network to get an accurate flux density estimation in known PS positions.

To test the robustness of PoSeIDoN, we apply it to slightly different situations without additional training. On the one hand, we apply it at 143 (top panel) and 353 GHz (bottom panel) with the same galactic cut ($|b| > 30^\circ$). The first channel has a lower Galactic emission but higher instrumental noise and bigger beam. The second one has the same beam as 217 GHz but higher instrumental noise and the Galactic thermal emission is stronger. On the other hand, the FCN is applied again at 217 GHz but allowing patches at lower galactic latitudes, $|b| > 10^\circ$, that implies stronger Galactic emission (see Fig. 4).

At 143 GHz (Fig. 3 top panel) the performance of PoSeIDoN with respect to the completeness is almost the same as the MHW2 with a $4\sigma$ threshold. The 90% completeness level in this channel is 224 mJy, 298 mJy and 300 mJy for the 3σ MHWF, the 4σ MHWF and PoSeIDoN, respectively. By comparing with the 217 GHz case, the FCN performance has worsen with respect to the MHW2 ones. The most probable reason is the change in the instrumental beam that will produce PS slightly bigger that the ones used to train (and thus expected by) the FCN at 217 GHz. This issue also explain PoSeIDoN under-estimation of the flux densities ($p_{\text{conf}} < 1$ for almost all detected PS). On the other hand, PoSeIDoN is still the most robust technique as for spurious PS. While both MHW2 results have already more than 20% spurious PS at the 90% completeness level, PoSeIDoN detects lower number of spurious PS well below $\sim 200$ mJy.

At higher frequencies the Galactic emission is stronger and the spurious PS issue is much worse. This is clearly shown in the spurious results of all the techniques at 353 GHz (Fig. 4 lower panel). The level of spurious PS is always above 50% for both MHW2 cases. PoSeIDoN is performing slightly better, although such issue is still present. At this frequency, a more conservative Galactic masking is needed or additional steps are required to decrease the number of spurious PS (see Planck Collaboration et al. 2014b, 2016b). As for the completeness, the levels are similar to the 217 GHz case: 154, 227 and 227 mJy for the 3σ MHWF, the 4σ MHWF and PoSeIDoN, respectively. Again, as in the 143 GHz case, PoSeIDoN completeness results are equal to the 4σ MHWF. More or less the flux densities estimation results are also similar to the 217 GHz case.

So, for all three cases the source detection worsen with higher frequencies due to the increase of the foregrounds contribution to the total map, being PoSeIDoN the overall best performing method.

To complete the analysis of the robustness of PoSeIDoN, we perform an additional test at 217 GHz. Without any additional training, we apply the FCN to a new set of validation simulations at 217 GHz but using a less aggressive Galactic mask of $|b| > 10^\circ$ (see Fig. 4). The completeness and flux density estimation results remains more or less the same as in the 217 GHz case with $|b| > 30^\circ$ for all the techniques. The 90% completeness levels in this case are 135, 195 and 149 mJy for the 3σ MHWF, the 4σ MHWF and PoSeIDoN, respectively. However, the spurious PS numbers increase dramatically, similarly to the 353 GHz case. Again, PoSeIDoN gives better results with this issue and it shows a more linear increase of spurious PS with respect to the 353 GHz case. This difference can be an indication that, as expected, the more the situation resembles the simulations training set, the better is the performance of the FCN.

Therefore, by training PoSeIDoN in each particular situation the results can be slightly improved, although probably not much comparing with the MHW2. The detection of Galactic sources inside the complicate Galactic plane is most likely the most interesting case where the re-training would significantly improve the results. However, we have demonstrated that even without specialised training, a FCN is able to compete with the MHW2 filtering schema, when applied to typical CMB experiment observed patches.

Finally, in the completeness panels of Figures 3 and 4 we also point out, with the grey dotted line, the PCCS2 90% completeness flux density limit (Planck Collaboration et al. 2016b): 177, 152 and 304 mJy at 143, 217, 353 GHz, respectively. Such values are in fair agreement with our findings. However, it should be stressed that such information has been added to guide the reader and it is not meant as a direct comparison with our results. First of all it must be taken into account that the PCCS2 is built to ensure at least 80% reliability. Then, the different masking must also be considered. As for the percentage of masked sky, the PCCS2 excludes the 15%, 35.1% and 52.4% for 143, 217 and 353 GHz (Planck Collaboration et al. 2016b), whereas our 30° Galactic cut corresponds approximately to a 50% of the sky. Moreover, the PCCS2 masks are tailored to avoid the most contaminating Galactic areas and to maximise the sky coverage of the catalogue. We could have used a more effective masking, but for our comparison-between-techniques purposes a simple galactic cut is enough. This point should just be taken into account when comparing with PCC2 numbers.

5. Conclusions

In this work, we successfully apply PoSeIDoN to the detection of sources in a realistic situation: we include simulated PS and CIB at the Planck frequencies of 143, 217 and 353 GHz; we add CMB and Galactic emission by randomly choosing the patches in the real Planck CMB (provided by the SEVEM method) and the Galactic simulated (provided by the FFP10 simulations) maps; finally we also add the instrumental noise, according to the Planck characteristics.

The network was trained at 217 GHz with a Galactic cut of 30°, using 50000 simulations. Then it was applied to the validation simulations at 143, 217 and 353 GHz and Galactic cut of 30°. At 217 GHz the network was also tested with a Galactic cut of 10°. Such results were then compared with those coming from the application of the MHW2 technique: in the overall, PoSeIDoN is performing better, providing more reliable results at lower flux densities.
In should be stressed that in the MH2 case, in order to get rid of the many spurious sources detected at low fluxes, we need to increase the flux density detection limit from 3σ to 4σ. On the contrary, PoSeIDoN application is straightforward, well performing even at 1σ (i.e. the results given in this work).

Another advantage of PoSeIDoN with respect to MH2 is that it doesn’t have border effects like any filtering approach. In the MH2 analysis we need to remove those pixels near the patch border, subsequently missing those sources falling in that region (they can be recovered by selecting overlapping patches as done in Planck Collaboration et al. 2011a, 2014b, 2016b). PoSeIDoN is not affected by such problem, being able to detect sources placed near the patch limits.

As expected, both methods worsen their performance with increasing frequencies, i.e. with the increase of the relative importance of the contaminants (mainly due to the Galactic thermal emission in our set of simulations). Moreover, they also get worse with a smaller galactic cut because, as expected, the Galactic contamination is higher.

As a reference, we also indicate the flux density limit at 90% completeness for the PCCS2, which is in fair agreement with our results. However it should be kept in mind that the PCCS2 is built imposing an overall 80% of reliability and by using tailored sky mask to better avoid Galactic contamination and preserving as much sky coverage as possible.

Finally, as a limit of PoSeIDoN, it must be said that the flux density estimation of the FCN method is not optimal, at least with respect to the MH2: the network behaviour in flux density estimation is to give lower flux densities with respect to the true ones. We notice a trend to assign lower flux densities to less reliable sources. So, our advice when building a catalogue (which is beyond the scope of this work) is to first blindly detect sources in a map with PoSeIDoN and then estimate the flux density of the retrieved sources by non-blindly applying some flux density estimation methods (e.g. non-blind MH2) in the obtained PoSeIDoN positions. A future development of the current work would be to train a second neural network to derive accurate flux density estimations in known PS positions.

Acknowledgements. LB and JGN acknowledge financial support from the PGC 2018 project PGC2018-101948-B-I00 (MICINN, FEDER) and PAPI-19-EMERG-11 (Universidad de Oviedo). JGN acknowledges financial support from the Spanish MINECO for the ‘Ramon y Cajal’ fellowship (RYC-2013-13256). JGC acknowledges financial support from the Spanish MINECO for the ‘Ramon y Cajal’ fellowship (RYC-2013-13256). JGC also acknowledges financial support from the European Union’s Horizon 2020 research and innovation programme under the H2020-INFRAIA-2018-2020 grant agreement No 101048962. This research has made use of the python packages Lythropy (Pérez & Granger 2007), matplotlib (Hunter 2007) and Scipy (Jones et al. 2001).

References
Barreiro, R. B., Sanz, J. L., Herranz, D., & Martínez-González, E. 2003, MNRAS, 342, 119
Bennett, C. L., Bay, M., Halpern, M., et al. 2003, ApJ, 583, 1
Blain, A. W., Ivison, R. J., & Smail, I. 1998, MNRAS, 296, L29
Cai, Z.-Y., Lapi, A., Xia, J.-Q., et al. 2013, ApJ, 768, 21
Cayón, L., Sanz, J. L., Barreiro, R. B., et al. 2000, MNRAS, 315, 757
Chauvin, Y. & Rumelhart, D. E. 2013, Backpropagation: theory, architectures, and applications (Psychology Press).
Dai, J., Li, Y., He, K., & Sun, J. 2016, in Advances in neural information processing systems, 379–387
Delabrouille, J., de Bernardis, P., Bouchet, F. R., et al. 2018, J Cosmology Astropart. Phys., 2018, 014
Dole, H., Lagache, G., et al. 2006, A&A, 451, 417
Dole, H., Rieke, G. H., Lagache, G., et al. 2004, ApJS, 154, 93
Fernández-Cobos, R., Viola, P., Barreiro, R. B., & Martínez-González, E. 2012, MNRAS, 420, 2162
Fluri, J., Kacprzak, T., Lucchi, A., et al. 2019, Phys. Rev. D, 100, 063514
Giusarma, E., Reyes Hurtado, M., Villascsua-Navarro, F. et al. 2019, arXiv e-prints, arXiv:1910.08555
Gómez, S. L. S., González-Gutiérrez, C., Alonso, E. D., et al. 2019, Publications of the Astronomical Society of the Pacific, 131, 108012
González-Nuevo, J., Argüeso, F., López-Caniego, M., et al. 2006, MNRAS, 369, 1603
González-Nuevo, J., Massardi, M., Argüeso, F. et al. 2008, MNRAS, 384, 711
González-Nuevo, J., Toftolatti, L., & Argüeso, F. 2005, ApJ, 621, 1
Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759
Granato, G. L., De Zotti, G., Silva, L., Bressan, A., & Danese, L. 2004, ApJ, 600, 580
Graves, A., Mohamed, A.-r., & Hinton, G. 2013, in 2013 IEEE international conference on acoustics, speech and signal processing, IEEE, 6645–6649
Hanany, S., Alvarez, M., Artis, E., et al. 2019, arXiv e-prints, arXiv:1902.10541
Hauzer, M. G. & Dwek, E. 2001, ARA&A, 39, 249
He, S., Li, Y., Feng, Y., et al. 2019, Proceedings of the National Academy of Science, 116, 13825
Herranz, D., Sanz, J. L., Barreiro, R. B., & Martínez-González, E. 2002, A&A, 395
Herranz, D., Sanz, J. L., Barreiro, R. B., & Martínez-González, E. 2002, A&A, 395
Hernández, L. B., Argüeso, F., López-Caniego, M., et al. 2006, MNRAS, 370, 932, 733
Mathuraya, A., Bard, D., Mendiylgraf, P., et al. 2018, arXiv e-prints, arXiv:1808.04728
Matsumura, T., Akiba, Y., Borrill, J., et al. 2014, Journal of Low Temperature Physics, 176, 733
Mustafa, M., Bard, D., Bhimji, W., et al. 2019, Computational Astrophysics and Cosmology, 6, 1
Nair, V. & Hinton, G. E. 2010, in Proceedings of the 27th international conference on machine learning (ICML-10), 807–814
Pearson, W. J., Wang, L., Alpaslan, M., et al. 2019, A&A, 631, A51
Pérez, F. & Granger, B. E. 2007, Software for Exa-Scale Simulations, 9, 21
Pérez, F. & Granger, B. E. 2007, Software for Exa-Scale Simulations, 9, 21
Pérez, F. & Granger, B. E. 2007, Software for Exa-Scale Simulations, 9, 21
Pérez, F. & Granger, B. E. 2007, Software for Exa-Scale Simulations, 9, 21
Pérez, F. & Granger, B. E. 2007, Software for Exa-Scale Simulations, 9, 21
Pérez, F. & Granger, B. E. 2007, Software for Exa-Scale Simulations, 9, 21
Pérez, F. & Granger, B. E. 2007, Software for Exa-Scale Simulations, 9, 21
Pérez, F. & Granger, B. E. 2007, Software for Exa-Scale Simulations, 9, 21
Pérez, F. & Granger, B. E. 2007, Software for Exa-Scale Simulations, 9, 21
Pérez, F. & Granger, B. E. 2007, Software for Exa-Scale Simulations, 9, 21
Pérez, F. & Granger, B. E. 2007, Software for Exa-Scale Simulations, 9, 21
Pérez, F. & Granger, B. E. 2007, Software for Exa-Scale Simulations, 9, 21
Fig. 3. Validation results at 143 (top panels), 217 (middle panels) and 353 GHz (bottom panels). Left column: Completeness (top panel) and reliability (bottom panel) for 30° galactic cut. The red solid line refers to the results obtained with PoSeIDoN, the blue dashed one to the 3σ MHW2 and the dot-dashed cyan to the 4σ MHW2. The dotted grey vertical line is the 90% completeness flux density limit for the PCCS2 (Planck Collaboration et al. 2016b). Right column: Relative flux density comparison between MHW2 (3σ blue hatched histogram and 4σ cyan squared histogram) and PoSeIDoN (red filled histogram) results for 30° galactic cut.
Fig. 4. For the 217 GHz case and galactic cut at 10°: completeness and reliability (left figure, top and bottom panels respectively), relative flux density comparison (right figure). Red solid line (filled histogram) refers to PoSelDoN, blue dashed line (hatched histogram) to 3σ MHW2 and cyan dot-dashed line (squared histogram) to 4σ MHW2. The dotted gray line points to the flux density limit at 90% completeness for the PCCS2.