SF2A 2018
P. Di Matteo, F. Billebaud, F. Herpin, N. Lagarde, J.-B. Marquette, A. Robin, O. Venot (eds)

A CATALOG PIPELINE FOR SOURCES IN THE CTA GALACTIC PLANE SURVEY

J. Cardenzana*[1]

Abstract. The upcoming Cherenkov Telescope Array (CTA), designed to observe Cherenkov radiation from very-high energy gamma-rays incident on Earth's atmosphere, will provide a significant improvement in both sensitivity and angular resolution compared to current generation imaging atmospheric Cherenkov telescopes. A key science goal of CTA is a survey of the entire Galactic plane. Outcomes of this survey include a census of Galactic gamma-ray source populations (supernova remnants, pulsar wind nebulae, gamma-ray binaries, etc), identifying possible PeVatron candidates, characterizing the diffuse Galactic gamma-ray emission and improving our knowledge of the origin of cosmic rays. However, in order to exploit the data for these purposes, an understanding of the underlying sources present in the survey data will be necessary. The Galactic plane survey presents many challenges including disentangling sources from underlying diffuse emission from galactic cosmic-rays and source confusion. This proceedings will describe current efforts to develop a pipeline for cataloging sources in CTA data built with the *ctools* analysis software. Specific focus will be given to the algorithms used for source detection and characterization in the anticipated CTA Galactic plane survey.

Keywords: gamma-ray, galactic, survey, catalog

1 Introduction

The Cherenkov Telescope Array (CTA) will be the most sensitive imaging atmospheric Cherenkov telescope (IACT) constructed (The Cherenkov Telescope Array Consortium et al. 2017). CTA will consist of two arrays of IACTs covering an energy range of 20 GeV up to 300 TeV. At energies above 1 TeV CTA will achieve an angular resolution better than three arcminutes and an energy resolution better than 10%. Combined with other improvements, CTA will have a sensitivity of 5-20 times better than existing IACT detectors.

One of the key science programs for CTA will be a survey of the Galactic plane. This Galactic plane survey (GPS) will cover all Galactic longitudes and Galactic latitudes within 5° of the Galactic plane. A total observation time of 1620 hours (1020 hours by the southern array and 600 hours by the northern array) spread over 3270 pointings is planned. The survey will achieve a sensitivity to sources with at least 1.3×10^{-12} erg cm^{-2} s^{-1} integral flux (above 100 GeV) for point sources across the entirety of the Galactic plane. An increased sensitivity of 5.4×10^{-13} erg cm^{-2} s^{-1} will be obtained in the inner region of the Galaxy ($|l| < 60°$). This proceedings will describe efforts to develop a pipeline for extracting a catalog of sources from the CTA GPS.

2 Methods

The pipeline described here is based on the *ctools* analysis software (Knödlseder et al. 2016). The general method of the pipeline works by searching for new sources seeds (§2.1) and fitting these seeds to the data (§2.2). This process is repeated until the catalog is determined to have reasonably converged.

* on behalf of the CTA Consortium

[1] Institut de Recherche en Astrophysique et Plantologie, 9 avenue du Colonel Roche, 31028 Toulouse, Cedex 4, France

2.1 Source Seed Detection

Source detection is conducted on a binned significance map with pixel widths of 0.05°. The significance is computed using a Poisson likelihood comparing the observed counts (derived from observations) to predicted counts in each pixel. The predicted counts map is initially computed assuming emission is strictly background in origin while later searches also account for sources that have been found and fit in previous steps. Both the observed and predicted counts maps are smoothed by a Gaussian with σ=0.05°, consistent with the spatial pixel size of the maps.

The resulting significance map is searched via a technique based on the SExtractor algorithm (Bertin & Arnouts 1996). This involves looping over each pixel in the significance map. If a pixel has a significance above a given threshold ($\geq 3\sigma$ for the work presented in this proceedings) it is identified as an *image pixel*. Groupings of at least eight image pixels are identified as potential source objects. Each potential source object is then put through a *deblending* procedure to identify substructures. This deblending process increases the possibility of disentangling multiple overlapping sources. The final list of deblended source objects is then used as seed sources for the fitting procedure.

2.2 Source Fitting Procedure

Each source seed is independently fit in three dimensions consisting of right ascension, declination and energy via a binned likelihood analysis. The analysis region is binned using a spatial pixel width of 0.02°, smaller than the instrument PSF, and 20 logarithmically spaced energy bins from 100 GeV to 100 TeV. The spatial and spectral parameters of the source seed along with those of the background models are jointly fit. Spatial and spectral parameters of nearby sources that overlap this analysis region are fixed at their previously fitted values. Only those sources deemed marginally significant (i.e. $TS \geq 10$)* are kept in subsequent iterations.

Initially, source seeds from the detection step in §2.1 are represented as point sources with power-law (PL) spectra. After fitting the point source model, the source is tested for extension starting with a disk model. If the disk model is determined to be favorable to the point source model (i.e. TS_{disk} - $TS_{point} \geq 10$) then a Gaussian model is also tested. This conditional testing of the Gaussian model is done to reduce the overall computation time of the pipeline. The extended source model with the larger TS is chosen as the best fit spatial source model, so long as it satisfies the requirement that TS_{ext} - $TS_{point} \geq 10$. Source extension is also restricted to a maximum value of 3° to minimize the likelihood of incorrectly detecting background interstellar gamma-ray emission as a source.

The angular width and height of the analysis region is chosen according to Eq. 2.1, where *width* is measured in degrees. r_{source} represents either 1.5 times the radius of a disk spatial model, or five times the standard deviation in the case of a Gaussian spatial model. An additional 2° is added to provide a buffer for the PSF of the instrument and for slight spatial movements while fitting the source position.

$$width = 2 + 2 \times \begin{cases} r_{source} & r_{source} > 0.5^\circ \\ 0.5 & r_{source} \leq 0.5^\circ \end{cases} \tag{2.1}$$

After all sources in the pipeline have converged and their best fit spatial model determined, each source is then tested for spectral curvature. This involves testing both log-parabola and exponentially cutoff power-law spectral models. The curved spectral model with the largest TS value is chosen, provided TS_{curved} - $TS_{PL} \geq 10$.

Finally, all sources are refit to find the best fit global values. This is done to ensure each source model accounts for the best fit spatial and spectral parameters of all other sources.

3 Performance

3.1 Overview

The performance of the catalog pipeline was assessed using a simulation roughly approximating the expected CTA GPS, as described in §1. The background models used in the simulation consist of an instrumental

*Here, TS=-2 $log(L_0/L_1)$ where L_0 and L_1 are the likelihood values obtained from a fit excluding and including the source model (respectively) to the data.

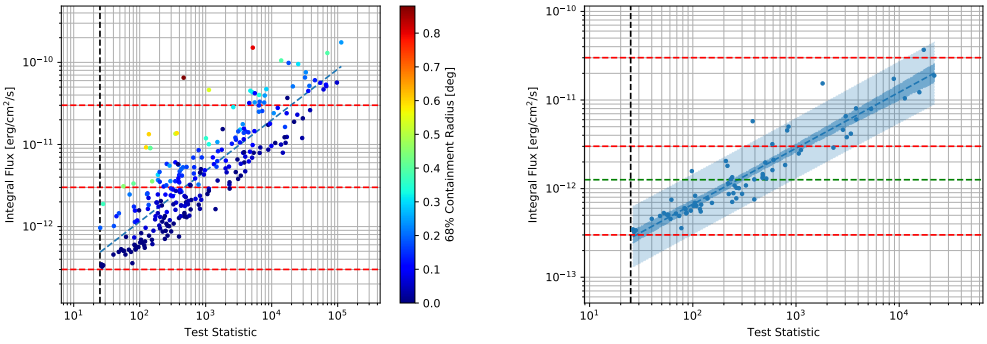

Fig. 1. Distribution of source *TS* vs. integral flux in the final catalog. Red lines denote integral fluxes equivalent to 10%, 1% and 0.1% of the Crab nebula. The black dashed line denotes *TS*=25, while the blue line shows the best fit to the data. **Left**: Distribution for all sources in the catalog. Points are colored based on their 68% containment radius. **Right**: Distribution for point sources only. The 95% confidence band on the fitted trend line (*dark blue*) and the 95% prediction band for sources detected in the catalog (*light blue*) are also shown. The green dashed line highlights the expected full-plane sensitivity of 1.3×10^{-12} erg cm^{-2} s^{-1} integral flux, above 100 GeV.

background model (approximating the residual background from interactions of diffuse cosmic-rays with the atmosphere) and a model for the TeV interstellar emission (IEM). For this first assessment of the technique, the IEM used in the simulation was also used in each of the source fits. This was done to ensure an accurate comparison of the resulting catalog sources could be made to those input to the simulation.

The analysis of the simulated survey only considers events between 100 GeV and 100 TeV. Note that although the simulation included an accounting for energy dispersion, this effect was ignored in the modeling for reasons related to computation time. It is expected that the lower-energy threshold of 100 GeV is sufficiently high to mitigate any significant impacts this will have on the results.

The simulated GPS was broken up into 18 individual regions to be processed in parallel. Each region was setup to cover a range of $|b| < 7.5°$ to allow extended sources near the edge of the analysis region to be properly accounted for. Each region also covered a range of 24° in Galactic longitude, yielding a 4° overlap with neighboring regions on either side. This overlap between regions is intended to ensure proper fitting of extended sources near the region boundaries. After the analysis of a given region is complete, only those sources located within $|b| < 5°$ and the 20° range in Galactic longitude associated with that region are kept.

3.2 Catalog Sensitivity

The distribution of *TS* vs. integral flux is shown in Fig. 1 for all sources in the catalog and for point sources only. Both plots demonstrate a trend in which brighter sources are detected with higher significance, as expected. The *left* plot also shows that more compact sources are detected with higher significance than less compact sources with the same integral flux.

The *right* plot in Fig. 1 includes the 95% prediction band (light blue band). 95% of detectable point sources should lie within this light blue band. Notably, this band passes below the planned sensitivity threshold (dashed green line) for the entire CTA GPS at *TS* ≈80. This suggests that CTA is likely to meet the target sensitivity with the planned survey.

3.3 Comparison with Simulation

The derived catalog can be directly compared to the input simulation. An association for a given catalog source is determined by finding the simulated source that is closest in position and extension. If no source is found whose centroid is within 0.05° and also having an extension within 0.2° of a given catalog source, then the source is considered *unassociated*. Since the accuracy of the derived integral source flux is of interest, integral flux is not used to establish association.

Using this criteria approximately 80.6% of the catalog sources are determined to have associations. The majority of the remaining 19.4% of sources are not false detections, but rather arise from large, extended sources

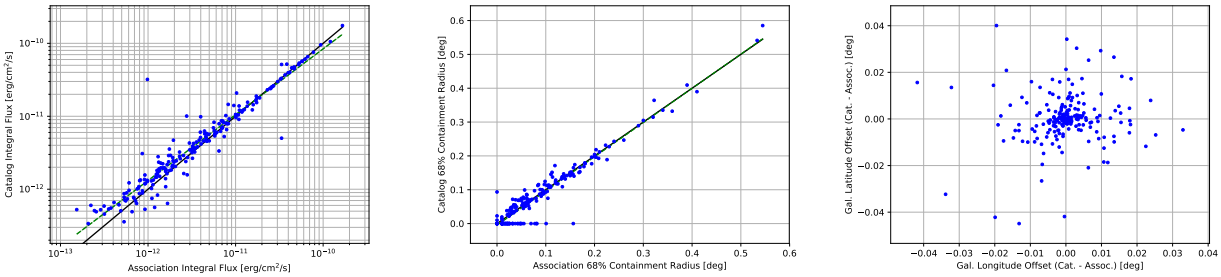

Fig. 2. Comparison of resulting catalog sources to the best association in the simulated list of sources. Plots show a comparison of integral flux (**left**), 68% containment radius (**middle**) and centroid offset (**right**). The green dashed line in the left and middle plots indicate the best fit line to the points. The black line represents a perfect reconstruction of the input simulation.

that cannot be correctly modelled as a single disk or Gaussian. In such cases the source detection algorithm picks out the residual emission as additional source seeds. These seeds are fit and ultimately determined to be significant due to the large residual flux of these extended sources.

A comparison of source characteristics is shown in Fig. 2. In general, the agreement between the catalog and associated simulated source is fairly good. Notably, there are two obvious outliers in the integral flux comparison plot (left plot, Fig. 2). Investigating these points indicates they are actually mis-associations. The point in the upper left is actually more closely associated with an extended nebula, but is closest in position with a pulsar that overlaps it. The outlier just right of the middle is actually a double association. This region was simulated as two nearly overlapping Gaussian models, however the fitting procedure was unable to establish an extension for one of these sources. The catalog ended up with a single Gaussian model trying to pickup the emission from the two extended sources while a single point source was found in the residuals near the brighter of the two extended sources.

4 Conclusions and Future Work

Preliminary studies using the *ctools* based catalog pipeline described here show that the pipeline is able to reasonably reconstruct the bulk of the input source characteristics. Still, this version of the pipeline is preliminary and has revealed at least two areas in which improvements could be investigated.

In the study presented here, source detection was done by first integrating over the full energy range of the simulated observations before computing the significance map. This could result in a bias towards detecting sources with softer spectra, as these sources will produce a higher number of overall photons for the same integral flux. Breaking up the source detection into a number of energy bins, computing the significance of each pixel in each energy bin and finally combining the significance from each energy bin for each pixel could potentially increase sensitivity to harder sources.

Source confusion is a notable problem in many catalogs. Currently each source is independently fit, which makes it more difficult for nearby sources (particularly those overlapping the source in question) to react to any changes in the spatial and/or spectral parameters. This can be especially problematic when establishing whether or not a source is extended. Instead, simultaneously fitting all sources within a given region of the sky would serve to mitigate this issue.

This work has been carried out thanks to the support of the OCEVU Labex (ANR-11-LABX-0060) and the A*MIDEX project (ANR-11-IDEX-0001-02) funded by the "Investissements d'Avenir" French government program managed by the ANR.

References

Bertin, E. & Arnouts, S. 1996, A&AS, 117, 393

Knödlseder, J., Mayer, M., Deil, C., et al. 2016, A&A, 593, A1

The Cherenkov Telescope Array Consortium, Acharya, B. S., Agudo, I., et al. 2017, ArXiv e-prints:1709.07997