

Implementation of a goodness-of-fit test for finding optimal concurrent radio and γ -ray pulsar light curves

A S Seyffert¹, C Venter¹, A K Harding², J Allison³ and W D Schutte³

¹ Centre for Space Research, North-West University, 11 Hoffman Street, Potchefstroom, 2531, South Africa

² Astrophysics Science Division, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA

³ School for Computer, Statistical, and Mathematical Sciences, North-West University, 11 Hoffman Street, Potchefstroom, 2531, South Africa

E-mail: 20126999@nwu.ac.za

Abstract. Since the launch of the *Fermi* Large Area Telescope in 2008 the number of known γ -ray pulsars has increased immensely to over 200, many of which are also visible in the radio and X-ray bands. Seyffert et al. (2011) demonstrated how constraints on the viewing geometries of some of these pulsars could be obtained by comparing their observed radio and γ -ray light curves by eye to light curves from geometric models. While these constraints compare reasonably well with those yielded by more rigorous single-wavelength approaches, they are still a somewhat subjective representation of how well the models reproduce the observed radio and γ -ray light curves. Constructing a more rigorous approach is, however, made difficult by the large uncertainties associated with the γ -ray light curves as compared to those associated with the radio light curves. Naively applying a χ^2 -like goodness-of-fit test to both bands invariably results in constraints dictated by the radio light curves. A number of approaches have been proposed to address this issue. In this paper we investigate these approaches and evaluate the results they yield. Based on what we learn, we implement our own version of a goodness-of-fit test, which we then use to investigate the behaviour of the geometric models in multi-dimensional phase space.

1. Introduction

The *Fermi* Large Area Telescope (LAT) is the primary instrument on the *Fermi Gamma-ray Space Telescope* mission which was launched on 11 June 2008. To date *Fermi* LAT has detected more than 200 new γ -ray pulsars, a dramatic increase over the 6 or 7 known at the time of *Fermi* LAT's launch [1]. This sudden abundance of pulsars visible in both the radio and γ -ray bands has opened the door to multiwavelength investigations that were not possible before.

These studies have, however, been made difficult by the comparatively low γ -ray photon flux (typically), as compared to the radio photon flux for any particular pulsar. This disparity manifests itself in the relative uncertainties on the observed intensities in the two wavebands, making traditional goodness-of-fit techniques very ineffective when trying to fit the radio and γ -ray light curves (LCs) simultaneously. This has led some investigators to adopt so-called by-eye methods to find optimal simultaneous fits (e.g., [2, 3]). A number of attempts have, however,

been made to address this issue directly, and this paper presents our latest such attempt. The most notable attempts to date ([4, 5]) artificially inflate the uncertainties accompanying the radio LCs in an attempt to grant the radio and γ -ray data equal weight in the determination of the optimal fit. As we will show, neither of their approaches accomplish this fully. We will then describe our proposed solution using these two approaches as reference.

2. The geometric pulsar models

We use an idealized picture of the pulsar system, wherein the magnetic field has a retarded dipole structure [6]. The γ -ray emission originates in regions of the magnetosphere (referred to as ‘gaps’) where the local charge density is sufficiently lower than the Goldreich-Julian charge density [7]. These gaps facilitate particle acceleration and consequently radiation. We assume that there are constant-emissivity emission layers embedded within the gaps, in the pulsar’s corotating frame. The location and geometry of these emission layers determine the shape of the γ -ray LCs, and there exist multiple models describing their geometry. In this paper only the two-pole caustic (TPC, [8]) γ -ray model will be examined. For the radio emission we use a simple empirical hollow cone model [9]. For a more detailed description of these models, see our previous work [3].

Fig. 1 (left) shows an example of a phaseplot produced using the TPC model. It is an equirectangular projection skymap of the emission of the pulsar, with rotational phase ϕ on the horizontal axis, and the observer’s line of sight ζ on the vertical axis. For each point in (ϕ, ζ) space it gives the relative intensity per solid angle of the emission that would be observed if the observer’s line of sight would pass through that point as the pulsar rotates. A phaseplot thus contains the projected all-sky γ -ray emission of a pulsar over one complete rotation.

The equator of the neutron star is at $\zeta = 90^\circ$, and its magnetic poles are at $\zeta = \alpha$ and $\zeta = 180^\circ - \alpha$. Thus, each value of α corresponds to a unique phaseplot. Fig. 1 shows an example of how a predicted LC can be obtained by cutting a phaseplot at a constant value for ζ (e.g., for a pulsar with $\alpha = 50^\circ$). The predicted LCs are independently normalised so their respective maxima are 1.

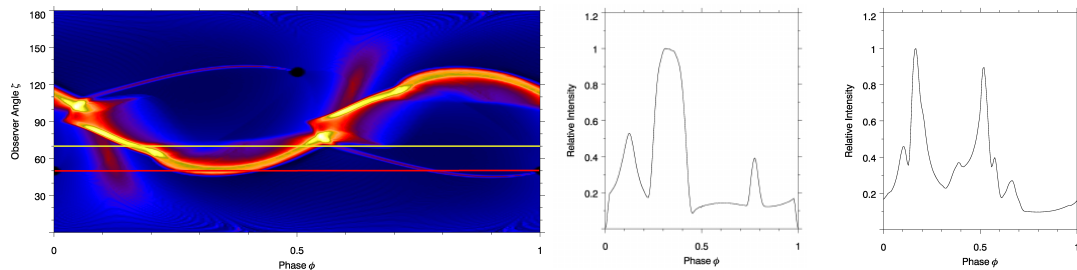


Figure 1. An example of constant- ζ cuts through a phaseplot for $\alpha = 50^\circ$. The cuts correspond to values of $\zeta = 50^\circ$ (red) and $\zeta = 70^\circ$ (yellow). We see that the caustics on the phaseplot present as peaks in the LC.

3. The fitting methods

3.1. The by-eye method

The by-eye method considers the LCs associated with the entire domain of α and ζ by generating so-called atlases. Each atlas consists of a set of predicted γ -ray and radio LCs generated at each point on a grid of (α, ζ) values covering the (α, ζ) domain at a chosen resolution. The data LCs are then compared to the model LCs at each point in this grid, and a subjective decision is made regarding the quality of each fit. Once the entire space has been scanned, a best-fit

contour emerges, and constraints on the viewing geometry (α, ζ) of the pulsar can be derived. An important aspect to note is that not only are the shapes of the LCs reproduced, but also the radio-to- γ phase lag. For an example of this method's application, see our previous study [3].

The by-eye method is not very rigorous, and would possibly yield different results when applied by different people. The results, however, do seem to be a good first attempt at finding concurrent fits when considered alongside other fits obtained independently ([10], [3]). Furthermore, this method allows us to discern which features of the LC most likely contribute to a specific LC's rejection, which in turn gives us insight into what effects are responsible for the shape of the final solution contours. Since the statistical methods aim to fit the same models (in our case at least) to the same data, we can use the knowledge garnered from the by-eye contours as a sanity check when considering the contours produced by the statistical methods.

3.2. The statistical methods

The statistical methods developed to date implement a modified version of Pearson's χ^2 goodness-of-fit test, with the test statistic defined as

$$X = \sum_{i=1}^{n_{\text{bins}}} \left(\frac{E_i - O_i}{U_i} \right)^2, \quad (1)$$

with E_i , O_i , and U_i corresponding to the model (expected) intensity, observed intensity, and error (uncertainty) on the observed intensity for the i th bin (of n_{bins} bins) [11]. This test statistic is χ^2 distributed with $N = n_{\text{bins}} - n_{\text{parameters}} - 1$ degrees of freedom. Ideally the minimum value obtained for X (as a function of the tested model's parameters) will be roughly N , indicating that there is some set of parameters for which the model fits the data well. Since the current models are still only approximate representations of the actual phenomena, this unfortunately rarely occurs, with the best fits having values for $X_{\text{min}} \gg N$. To compensate for this, and to allow constraints to be derived from our imperfect models, the value of X is normalised such that its minimum is N .

This method has been applied successfully for LCs from a single waveband, but has proven to be less useful when trying to fit observed LCs from multiple wavebands concurrently. Calculating the values of X for the LCs in each waveband and naively combining them (with equal weighting) for each set of parameter values, yields constraints that tend to be dominated by the waveband that has lower relative uncertainties. In our case, the radio goodness-of-fit test statistic dominates the overall fit.

This problem has been addressed in the methods developed thus far, but with limited success. One approach is to somehow artificially inflate the radio uncertainties (e.g., [4, 5]). How this is done differs depending on the method, but the results are invariably an improvement on a naive goodness-of-fit method. Fig. 2 shows how these artificially inflated uncertainties look (specifically the method used by [4]), and how the resulting values of X change.

The motivation behind these statistical methods is that for us to be able to add the χ^2 profiles for the radio and γ -ray goodness-of-fit tests, the values for X associated with the best radio and best γ -ray fit should be of roughly comparable size. This is best accomplished by adjusting the uncertainties of the radio LC. These methods do not work effortlessly for all pulsars, though. For some pulsars, one needs to increase the degree to which the radio uncertainties are inflated, while for others a decrease is required [4]. This seems to indicate that there might still be room for improvement.

4. Our proposed improvement

The method we propose takes a more direct route toward obtaining comparable radio and γ -ray values for X . In considering the statistical methods above we realised that ensuring comparable

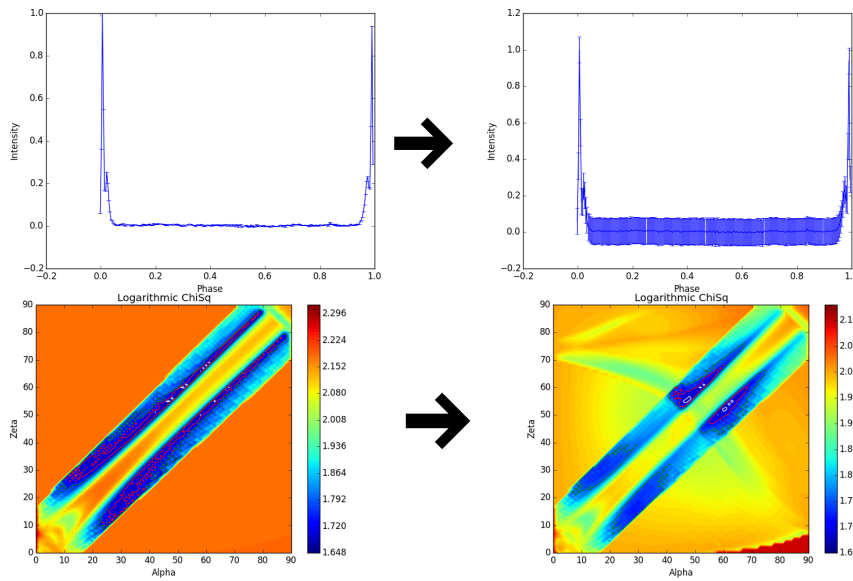


Figure 2. The top set of figures shows an example of radio uncertainties artificially inflated in a manner similar to that proposed by [4]. The bottom set of figures shows how this influences the resulting values for X for the combined radio and γ -ray fits. The left-most of the χ^2 maps finds its best fits along the diagonal, which means that the radio fit dominates. In the right-most χ^2 map we see the cross-diagonal features of the γ -ray fit emerging. The resulting contours become more constraining as both models start contributing more equally to the fitting process.

minima for the χ^2 maps is not sufficient as the minima will generally not be colocated in (α, ζ) space. Stated more clearly, we realised that the minimum of the γ -ray test statistic should be compatible with all the values of the radio test statistic, not just its minimum, and vice versa. This ensures that the minimum γ -ray test statistic value is not averaged with a (non-minimum) radio test statistic value that is orders of magnitude larger, when combining the radio and γ -ray χ^2 maps, since the radio will still dominate in that case.

To accomplish this compatibility we scale the radio test statistic values such that their dynamic range (the difference between their maximum and minimum) is equal to that of the γ -ray test statistic values. In effect we force the radio and γ -ray test statistic values to carry equal weight in the determination of the best concurrent fit. This scaling is done globally over the space defined by the parameters for which we are deriving constraints.

The procedure employed proceeds as follows:

- (i) **Calculate the unscaled χ^2 maps, using Eq. (1), for the radio and γ -ray fits independently.** In our case each model LC has four parameters associated with it, which means these maps will have four dimensions. The LC parameters are α , ζ , ϕ_0 (zero point in phase), and A (the LC amplitude).
- (ii) **Minimise each map over all parameters that are independent for the two wavebands.** For the concurrent fits in this study only A can be minimised over in this step. Since we assume that the radio and γ -ray emission originates from different regions in the same magnetic field, the value of ϕ_0 cannot be varied independently, and can therefore not be minimised over in this step. The resulting maps in our case have three dimensions.
- (iii) **Scale the γ -ray map to have a minimum equal to the appropriate number of degrees of freedom [5].** For simplicity, we assume here that this scaled test statistic is χ^2 distributed.

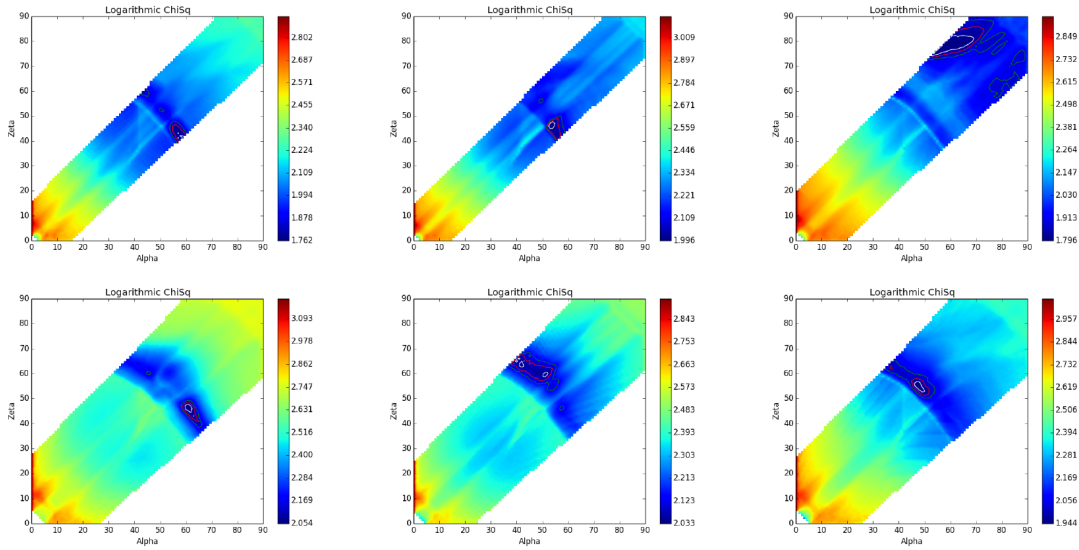


Figure 3. The resulting combined χ^2 maps for the six pulsars studied in our previous work [3]. The white, red, and green contour lines represent 1σ -, 2σ -, and 3σ -like contours respectively, but should be considered cautiously. The white regions are where the radio model predicts no LC.

- (iv) **Scale the radio map such that it has the same dynamic range as the freshly-scaled γ -ray map.** The scaled radio test statistic is also assumed to be χ^2 distributed, in accordance with the assumption made for the scaled γ -ray test statistic.
- (v) **Sum the two maps.** The resulting combined map will have as many dimensions as the models have coupled parameters (parameters that cannot be varied independently). In our case the coupled parameters are α , ζ , and ϕ_0 . The number of degrees of freedom for this combined test statistic is the sum of the minima of the constituent maps [12]. This follows from the assumption made in (iii) and (iv) for the two constituent maps.
- (vi) **Minimise the combined map over all coupled parameters except α and ζ .** For our study, and more generally, these two parameters are the ones for which constraints are to be derived. The resulting map therefore has two dimensions.
- (vii) **Obtain the confidence contours using the 1σ , 2σ , and 3σ values of a χ^2 distribution with 2 degrees of freedom.** Since the lowest value the minimum of the combined map can have is the sum of the minima of the constituent maps (in the case of colocated minima), we can derive the contours by adding the above-mentioned 1σ , 2σ , and 3σ values to this lowest minimum [12].

5. Results

Implementing this approach for the six pulsars we previously studied yields constraints that compare very well to those obtained by eye [3]. The inclusion contours shown were derived as discussed in step (vii) in the previous section. They should, however, only serve as an estimation of how the actual contours may look since it is not yet clear how the actual contours should be computed.

6. Conclusion and Future Work

The inclusion contours obtained using our proposed method compare favourably with those for the by-eye method, which in turn compare well with independently derived constraints [3].

Although the the aim of these statistical methods is not to reproduce the results obtained using the by-eye method, the results of the by-eye method serve as a sanity check which the statistical methods should pass.

The χ^2 maps produced using our method have a number of properties which may prove useful in future studies. Chief among these properties is that the contours are calculated relative to a value other than the minimum of the combined map. This allows our method to reject an unsuccessful model combination even though some normalisation is done in obtaining the combined map. Another useful property is that the combined map remains essentially unchanged if step (iii) is omitted, effectively only being scaled. This allows for the possibility of comparing the goodness of fit of different model combinations, and therefore identifying the most successful model combination.

The interplay between our confidence in the geometric models and the uncertainties associated with the observed LCs also seems to be an important one to understand. It is useful to differentiate between evaluating which predicted LCs fit *best*, and whether or not the observed LCs are fit *well*. For a single model, determining the goodness of fit for each possible solution not only tells us how *well* the data is fit, but also allows us to determine which of those solutions is the *best* fit. This correlation does not hold when we generalise our method to include multiple models and multiple datasets. The breakdown occurs because we calculate the combined test statistic as the sum of the two constituent test statistics, making it sensitive to changes in the scaling of either of the single-model test statistics. Since our confidence in each model could be expressed as a scaling of the test statistic, effectively weighting their contribution to the combined test statistic, it is difficult to ensure that the single-model test statistics are properly scaled before summation. Our approach to the scaling of the individual test statistics is an attempt to address this problem directly.

The most important open issue in this study is how the confidence contours are to be defined. The specification given in Sec. 4, based on the assumptions regarding the distributions associated with the radio and γ -ray test statistics, still needs to be tested. We are currently developing a Monte Carlo simulation strategy aimed at doing this. The method derived here will also be used to compare a variety of multi-model combinations based on concurrent fitting of pulsar LCs.

Acknowledgments

This work is based on research supported by the National Research Foundation (NRF) of South Africa (Grant Numbers 90822 and 93278). Any opinions, findings, and conclusions or recommendations expressed are that of the authors, and the NRF accepts no liability whatsoever in this regard. A.K.H. acknowledges support from the NASA Astrophysics Theory Program. C.V. and A.K.H. acknowledge support from the *Fermi* Guest Investigator Program.

References

- [1] Thompson D J 2001 *AIP Conf. Series* **558** 103–14
- [2] Venter C, Harding A K and Guillemot L 2009 *Astrophys. J.* **707** 800
- [3] Seyffert A S 2011, Venter C, Harding A K, Johnson T J *Proc. of The 56th Ann. SAIP Conf.* 531–36
- [4] Johnson T J et al. 2014 *Astrophys. J. Suppl. S.* **213** 1–54
- [5] Pierbattista M et al. 2015 *Astron. Astrophys.* **575** 1–88
- [6] Deutsch A J 1955 *Ann. d’Astrophys.* **18** 1–10
- [7] Goldreich P and Julian W H 1969 *Astrophys. J.* **157** 869–80
- [8] Dyks J and Rudak B 2003 *Astrophys. J.* **598** 1201–6
- [9] Story S A, Gonthier P L and Harding A K 2007 *Astrophys. J.* **671** 713–726
- [10] Weltevrede P et al. 2010 *Astrophys. J.* **708** 1426–41
- [11] Avni Y 1976 *Astrophys. J.* **210** 642–646
- [12] Lampton M, Margon B and Bowyer S 1976 *Astrophys. J.* **208** 177–90