

A Bayesian analysis of the 69 highest energy cosmic rays detected by the Pierre Auger Observatory

Alexander Khanin¹^{*} and Daniel J. Mortlock^{1,2}

¹*Astrophysics Group, Imperial College London, Blackett Laboratory, Prince Consort Road, London SW7 2AZ, UK*

²*Department of Mathematics, Imperial College London, London SW7 2AZ, UK*

Accepted 2016 May 12. Received 2016 April 12; in original form 2016 January 9

ABSTRACT

The origins of ultrahigh energy cosmic rays (UHECRs) remain an open question. Several attempts have been made to cross-correlate the arrival directions of the UHECRs with catalogues of potential sources, but no definite conclusion has been reached. We report a Bayesian analysis of the 69 events, from the Pierre Auger Observatory (PAO), that aims to determine the fraction of the UHECRs that originate from known AGNs in the Veron-Cety & Verson (VCV) catalogue, as well as AGNs detected with the *Swift* Burst Alert Telescope (*Swift*-BAT), galaxies from the 2MASS Redshift Survey (2MRS), and an additional volume-limited sample of 17 nearby AGNs. The study makes use of a multilevel Bayesian model of UHECR injection, propagation and detection. We find that for reasonable ranges of prior parameters the Bayes factors disfavour a purely isotropic model. For fiducial values of the model parameters, we report 68 per cent credible intervals for the fraction of source originating UHECRs of $0.09^{+0.05}_{-0.04}$, $0.25^{+0.09}_{-0.08}$, $0.24^{+0.12}_{-0.10}$, and $0.08^{+0.04}_{-0.03}$ for the VCV, *Swift*-BAT and 2MRS catalogues, and the sample of 17 AGNs, respectively.

Key words: methods: statistical – cosmic rays.

1 INTRODUCTION

Cosmic rays (CRs) are highly accelerated protons and atomic nuclei, some of which enter the Solar system and reach the Earth. They are the most energetic particles observed in nature, with energies in the range 10^9 – 10^{21} eV (see e.g. Kotera & Olinto 2011; Letessier-Selvon & Stanev 2011, for reviews).

A number of open scientific issues remain with respect to CRs, in particular ultrahigh energy cosmic rays (UHECRs) with arrival energies $E_{\text{arr}} \gtrsim 10^{19}$ eV. The study of UHECRs is complicated by the fact that they experience an abrupt cutoff in their energy spectrum at $\sim 4 \times 10^{19}$ eV, so that only small samples are available. The largest currently available sample is the 69 events with $E_{\text{arr}} \geq 5.5 \times 10^{19}$ eV recorded by the Pierre Auger Observatory (PAO) between 2004 January 1 and 2009 December 31 (Abreu et al. 2010).

One open issue in the study of UHECRs is the question of their sources. A number of candidates, such as active galactic nuclei (AGNs) and pulsars, have been proposed, but studies have not been conclusive (see e.g. Kalmykov et al. 2013 for a review). The question of UHECR origins can be studied by attempting to associate the arrival directions with their sources. While UHECRs are charged particles and therefore experience magnetic deflection as they propagate, they are sufficiently energetic that the total deflection is ex-

pected to be ~ 2 – 10° (e.g. Medina Tanco, de Gouveia Dal Pino & Horvath 1998; Sigl, Miniati & Enßlin 2004; Dolag et al. 2005), so that some information about their points of origin should be retained.

Association of UHECRs with catalogues of potential sources is made possible by the fact that UHECRs with energies of $E \gtrsim 5 \times 10^{19}$ eV are expected to have come from a limited radius of ~ 100 Mpc. This radius is sometimes called the Greisen–Zatsepin–Kuzmin (GZK) horizon and arises due to the fact that UHECRs at those energies scatter off the cosmic microwave background (CMB) radiation in a process known as the GZK effect (Greisen 1966, Zatsepin & Kuzmin 1966). The mean free path of the GZK effect at high energies is a few Mpc, and the energy loss in each collision is 20–50 per cent. The resultant attenuation is very rapid and is the cause of the cutoff in the UHECR energy spectrum observed by both HiRes (Abbasi et al. 2008b) and PAO (Abraham et al. 2008b).

A number of attempts have been made to find correlations between UHECR arrival directions and catalogues of possible sources. Cross-correlation studies have been conducted with galaxy catalogues, such as the Two Micron All-Sky Survey (2MASS) Redshift Survey (2MRS) (Abraham et al. 2009; Abbasi et al. 2010a), as well as specific types of objects such as AGNs (Abraham et al. 2007, 2008a; George et al. 2008; Pe’Er, Murase & Mészáros 2009; Watson, Mortlock & Jaffe 2011) and BL Lacertae objects (BL LACs; Tinyakov & Tkachev 2001). Overall, no clear consensus has been reached. Different studies have reported different degrees of

* E-mail: ak2008@imperial.ac.uk

correlation, depending on the statistical approach, the UHECR sample, and the population of source candidates that was used. The most significant correlation was reported by the Pierre Auger Collaboration, between arrival directions of UHECRs with energies $E \geq 5.7 \times 10^{19}$ eV and the positions of nearby AGNs (Abraham et al. 2007). The result was supported by Yakutsk data (Ivanov 2009), but not by HiRes (Abbasi et al. 2008a), or the Telescope Array (Abu-Zayyad et al. 2012). A more recent analysis of a larger PAO sample has shown a weaker correlation than before (Abreu et al. 2010).

The lack of consensus on these issues is partly due to the difficulty of analysing such small sample sizes. Given the small size of the UHECR data sets, it is important to utilize as much of the available information as possible. This can be achieved by adopting a Bayesian methodology that involves models of the relevant physical processes. The first steps to such a comprehensive Bayesian work have been made in the recent work of Watson et al. (2011) and Soiaporn et al. (2013).

Watson et al. (2011) analysed the 27 events that were analysed in Abraham et al. (2007), and derived a posterior for the fraction that originated from AGNs in the Veron-Cetty & Veron (VCV) catalogue (Véron-Cetty & Véron 2006). To do so, they used a two-component parametric model, characterized by a source rate Γ and a background UHECR rate R . The model assumed that the UHECR arrival directions are points drawn from a Poisson intensity distribution on the celestial sphere. The intensity distribution was obtained with a computational UHECR model. Watson et al. (2011) report strong evidence of a UHECR signal from the VCV AGNs. They find a low AGN fraction that is consistent with Abreu et al. (2010). For fiducial values of the model parameters, they report a 68 per cent credible interval for the AGN fraction of $F_{\text{AGN}} = 0.09^{+0.05}_{-0.04}$.

Soiaporn et al. (2013) developed a multilevel Bayesian framework to attempt to associate the 69 UHECRs that were recorded at the PAO in the period 2004–09 with 17 nearby AGNs catalogued by (Goulding et al. 2010, hereafter G10). They report evidence for a small but non-zero fraction of the UHECRs to have originated at the AGNs from G10, of the order of a few per cent to 20 per cent.

We extend the formalism of Watson et al. (2011) with both a greater data set and a refined UHECR model. Following Abreu et al. (2010), we extend the analysis to two further source catalogues: AGNs from the *Swift* Burst Alert Telescope (*Swift*-BAT; Baumgartner et al. 2010) and galaxies from 2MRS (Huchra et al. 2012). We also extend the analysis to the 17 AGNs from the G10 catalogue.

After discussing the UHECR and source data sets in Section 2, we explain our UHECR model in Section 3, discuss the statistical formalism of our Bayesian model comparison in Section 4, and the application of the formalism to mock data sets in Section 5. The results of applying the formalism to the PAO data are discussed in Section 6. Some aspects of our computational approach are described in Appendix A, and some subtleties of our model comparison are explored in Appendix B. We use a Hubble constant of $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, where required to convert between redshifts and distances.

2 DATA

2.1 UHECR sample

The sample of UHECR events that was used in this analysis were the 69 highest energy events recorded at the PAO between January 2004 and November 2009, as documented in Abreu et al. (2010).

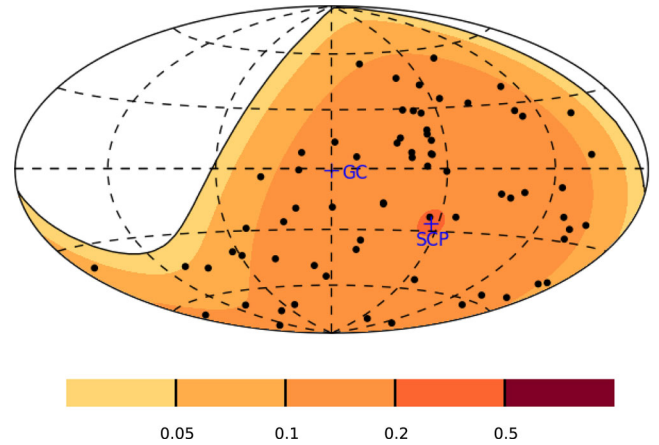


Figure 1. Relative PAO exposure in Galactic coordinates. The arrival directions of the 69 UHECRs are shown as black points. The Galactic Centre (GC) and south celestial pole (SCP) are indicated.

These are the events with observed energies E_{obs} above the threshold $E_{\text{obs}} \geq E_{\text{thres}} = 5.7 \times 10^{19}$ eV.

The PAO is a CR observatory located in Argentina, at a longitude of 69.5° W and a latitude 35.2° S. PAO is a hybrid observatory, which means that it uses both surface detection (SD) and fluorescent telescope detection (FD) of UHECRs. The observatory has SD plastic scintillators of a total area of 3000 km^2 and four FD telescope sites, with a total of 27 FD telescopes. The PAO's total exposure of this data set is $\epsilon_{\text{tot}} = 20370 \text{ km}^2 \text{ sr yr}$, and its relative exposure per unit solid angle, $d\epsilon/d\Omega$, is illustrated in Fig. 1. The relative exposure is directly proportional to $Pr(\text{det}|\mathbf{r})$, the probability that a UHECR will be detected if it arrives from direction \mathbf{r} , but is normalized so that $\int (d\epsilon/d\Omega) d\Omega = \epsilon_{\text{tot}}$. Fig. 2 shows the UHECR flux that would be expected if UHECRs were generated at the four source catalogues under consideration, modulated by PAO's relative exposure.

PAO measures UHECR arrival directions with an uncertainty of $\sim 1^\circ$ and arrival energies with a relative uncertainty of ~ 12 per cent (Letessier-Selvon et al. 2014).

2.2 Source catalogues

As potential source catalogues, we consider AGNs from the VCV, *Swift*-BAT and G10 catalogues, and galaxies from the 2MRS catalogue. This allows us to compare our analysis for the *Swift*-BAT and 2MRS sources with the analysis from Abreu et al. (2010), our analysis for the VCV sources with the analyses from both Abreu et al. (2010) and Watson et al. (2011), and our analysis of the G10 sources with Soiaporn et al. (2013).

We use the 12th edition of the VCV catalogue, selecting sources with $z_{\text{obs}} \leq 0.03$, as AGNs with higher redshift are too far away to be plausible UHECR sources, and can be shown to have a negligible effect on the results. We omit sources for which absolute magnitudes are not stated. The total number of VCV AGNs that meet those requirements is $N_{\text{VCV}} = 921$. This is the same sample of sources that was used in Abraham et al. (2007), Abreu et al. (2010) and Watson et al. (2011), and in PAO's more recent analysis Aab et al. (2015). While the VCV catalogue is heterogeneous and thus not ideal for statistical studies, it is close to complete for the low-redshift AGNs that are of relevance here.

For the *Swift*-BAT catalogue, we use the 58 month version that includes a total of $N_{\text{BAT}} = 1092$ sources. In the case of the 2MRS

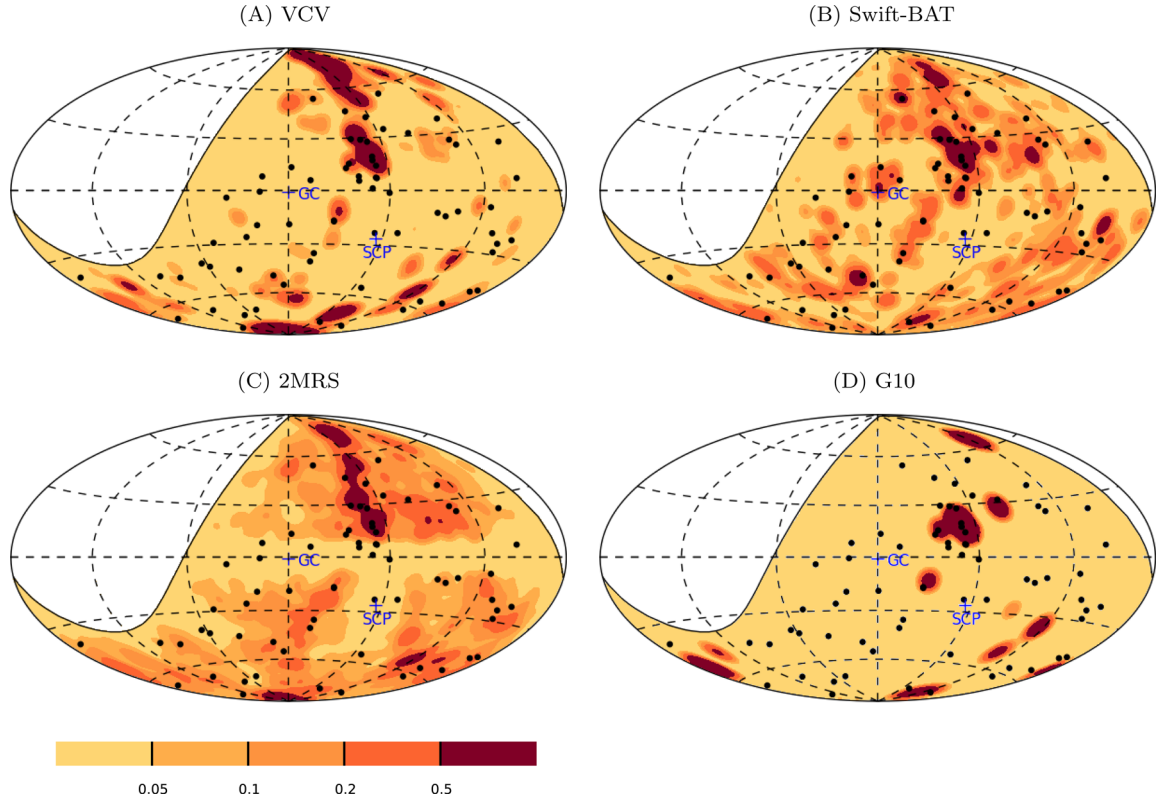


Figure 2. Positional dependence of the expected number of source originating events, for the VCV, *Swift*-BAT, 2MRS, and G10 catalogues. A fiducial value of the smearing parameter $\sigma = 3^\circ$ is assumed. The arrival directions of the 69 UHECRs are shown as black points. Galactic coordinates are used, and the GC and SCP are indicated.

catalogue, we used the catalogue version 2.4, 2011 Dec. 16. We exclude events that are within 10° of the Galactic plane, to avoid biases due to the incompleteness of the catalogue in the region of the Galactic plane. This leaves a total of $N_{2\text{MRS}} = 20\,702$ galaxies. These samples of *Swift*-BAT and 2MRS sources are the same as those used by Abreu et al. (2010).

The G10 catalogue is a well-characterized volume-limited sample of AGNs. The 17 AGNs contained in it constitute all infrared-bright AGNs within 15 Mpc. This is the same sample that was used by Soiaporn et al. (2013).

3 UHECR MODEL

A Bayesian UHECR analysis requires a realistic model of UHECR injection, propagation, and detection. This model was used both to compute the likelihoods in our statistical formalism (Section 4), and to create simulated mock catalogues of UHECRs to test our methods (Section 5). The model assumed a pure proton composition of UHECRs.

3.1 Injection

We adopt a model in which any given UHECR source emits UHECRs with an emission spectrum given by

$$dN_{\text{emit}}/dE_{\text{emit}} \propto E_{\text{emit}}^{-\gamma-1}, \quad (1)$$

where the logarithmic slope γ is taken to be 3.6 (Abraham et al. 2010). The spectrum is normalized in such a way that the total

emission rate of UHECRs with energy greater than E_{emit} is given by

$$\frac{dN_{\text{emit}}(> E_{\text{emit}})}{dt} = \Gamma \left(\frac{E_{\text{emit}}}{E_{\text{min}}} \right)^{-\gamma}, \quad (2)$$

where $E_{\text{min}} = 5.7 \times 10^{19}$ eV is the minimum UHECR emission energy, and Γ is the rate at which the source emits UHECRs with $E_{\text{emit}} > E_{\text{min}}$.

3.2 Energy loss during propagation

The energy loss processes experienced by UHECRs can be characterized in terms of the loss length $L_{\text{loss}} = -E(dE/dr)^{-1}$. Given the loss length as a function of energy, it is possible to calculate the total amount of energy that a UHECR loses as it travels to the Earth from a given distance by solving the differential equation

$$\frac{dE}{dr} = -\frac{E}{L_{\text{loss}}(E)}. \quad (3)$$

For pure proton composition, L_{loss} obeys the expression

$$L_{\text{loss}}^{-1} = \frac{1}{c} [\beta_{\text{GZK}}(E, z) + \beta_{\text{BH}}(E, z) + \beta_{\text{adi}}(E, z) + \beta_{\text{IR}}(E, z)], \quad (4)$$

where c is the speed of light and $\beta_{\text{GZK}}(E, z)$, $\beta_{\text{BH}}(E, z)$, $\beta_{\text{adi}}(E, z)$, and $\beta_{\text{IR}}(E, z)$ are terms corresponding to the main energy loss processes experienced by UHECRs of pure proton composition (e.g. Stanev 2009; Kotera & Olinto 2011):

(i) the photo-pion production due to the GZK scattering off the CMB photons at energies above $E \gtrsim 5 \times 10^{19}$ eV;

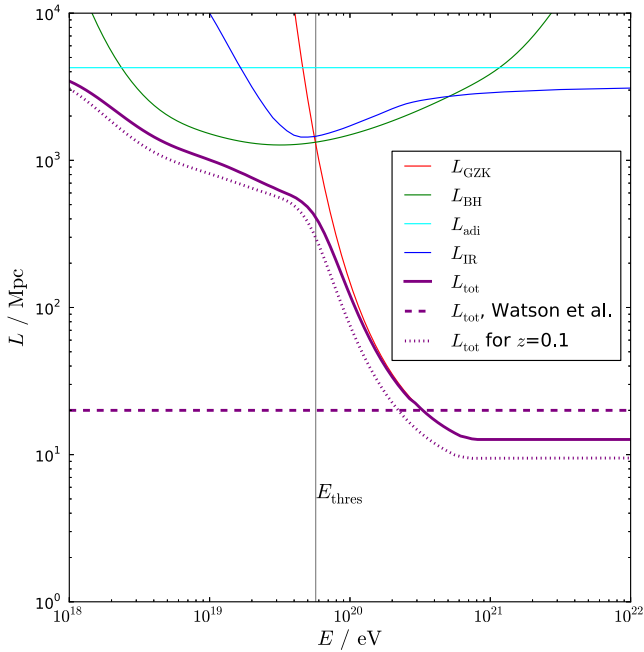


Figure 3. Loss lengths from the energy loss processes, compared to the constant loss length used by Watson et al. (2011), as described in Section 3.2.

(ii) Bethe–Heitler (BH) e^+e^- pair production (also a scattering process off the CMB radiation), which dominates at lower energies (Hillas 1968);

(iii) the adiabatic energy loss due to the expansion of the Universe;

(iv) the photopion production due to scattering off photons in the infrared-ultraviolet (IR-UV) spectrum.

A detailed discussion of the first three processes, including relevant mathematical expressions, can be found in De Domenico & Insolia (2013). For the pion production off the IR-UV photons, the relevant discussion can be found in Kotera & Olinto (2011). The loss lengths are shown as a function of energy in Fig. 3. The figure shows the energy loss lengths of the separate processes, as well as the combined total loss length L_{tot} . The figure shows L_{tot} plots for z values of 0.0 and 0.1, which correspond to distances of 0 and ~ 400 Mpc, thus, covering the GZK horizon. L_{GZK} appears very rapidly after an energy of $\sim 4 \times 10^{19}$ eV and begins to dominate the energy loss. As we are interested only in UHECRs with energies $E_{\text{obs}} > E_{\text{thres}} = 5.7 \times 10^{19}$ eV, the GZK scattering is the most relevant loss process in this investigation.

The energy dependence of $L_{\text{loss}}(E)$ is one of the main improvements of this propagation model over the model used in Watson et al. (2011), where L_{loss} was taken to be a constant. The constant value of L_{loss} used by Watson et al. (2011) is also displayed in Fig. 3 for comparison.

3.3 Effective smearing

We combined the magnetic deflection that a UHECR experiences during propagation and the uncertainty in its detected arrival direction into a single kernel, which was chosen to be a von Mises–Fisher (vMF) distribution, defined as

$$Pr(\hat{\mathbf{r}}|\hat{\mathbf{r}}_{\text{src}}, \kappa) = \frac{\kappa}{4\pi \sinh(\kappa)} \exp(\kappa \hat{\mathbf{r}} \cdot \hat{\mathbf{r}}_{\text{src}}), \quad (5)$$

where $\hat{\mathbf{r}}$ is the measured arrival direction of the ray, $\hat{\mathbf{r}}_{\text{src}}$ is the source direction, and κ is the concentration parameter. We refer to this as the buckshot model. The vMF distribution resembles a Gaussian on the sphere, with κ being inversely related to the width of the Gaussian: for large values of κ the distribution is peaked over an angular scale of $\sim 1/\sqrt{\kappa}$; if κ tends to 0 the distribution becomes uniform on the sphere.

The magnitude of the deflection that the highest energy UHECRs experience is uncertain, with the estimates of typical deflection angles ranging from ~ 2 to 10° (e.g. Medina Tanco et al. 1998; Sigl et al. 2004; Dolag et al. 2005). We assume a fiducial smearing angle of $\sigma \simeq 3^\circ$ ($\kappa = 360$), but also conduct investigations for smearing angles of $\sigma \simeq 6$ – 10° ($\kappa = 90$ and 30).

A limitation of the simple buckshot model is the fact that in the structured magnetic fields that are traversed by the UHECRs, and the deflection of the particles can be expected to be direction dependent. Models that take into account such direction dependence would yield correlated UHECR arrival directions. An example of an extension of the buckshot model that takes into account the direction dependence of the deflection has been discussed in Soiaporn et al. (2013), where the vMF distribution of a given source is centred not on the source itself, but on a particular guide direction that is associated with that source. Other studies have discussed and conducted simulations with more refined models of the magnetic fields that account for the direction dependence (e.g. Medina Tanco et al. 1998; Nagar & Matulich 2010; Farrar et al. 2013; Keivani, Farrar & Sutherland 2015). A possible extension of our work would be to incorporate such models into our framework, although it will not then be possible to use the counts-in-cells approach (Section 4.2) to evaluate the likelihood.

3.4 Observed UHECR flux

The number of UHECRs from source s above a threshold energy E_{thres} observed on Earth per unit area per unit time, $dN_s(E_{\text{obs}} \geq E_{\text{thres}})/dt dA$, is a quantity that is important in our statistical analysis. This rate is proportional to the rate of UHECRs emitted by the source, Γ_s , but it also depends on the distance dependence of the UHECR energy loss, and on the UHECR injection spectrum. We use the UHECR propagation model described in Section 3.2 to determine the injection energy corresponding to the threshold energy E_{thres} and to the source distance D_s . Combining this value with equation (2) and with the source distance D_s , we obtain

$$\frac{dN_s(E_{\text{obs}} \geq E_{\text{thres}})}{dt dA} = \frac{\Gamma_s}{4\pi D_s^2} \left[\frac{E_{\text{emit}}(E_{\text{thres}})}{E_{\text{min}}} \right]^{-\gamma}. \quad (6)$$

This expression assumes that the observed energy, E_{obs} , is equivalent to the arrival energy of the UHECR, E_{arr} . Thus, for the purposes of the calculation, the 12 per cent energy uncertainty of the PAO measurements is neglected. The variation in source rates, Γ_s , among the sources that we are considering is not negligible. We use the source rate of Centaurus A as the reference value Γ . The source rate of a source s is obtained by weighing the flux F_s of that source in a particular band against the flux F_{Cen} of Centaurus A in that same band. The waveband of the flux thereby is different depending on the source catalogue. For VCV, the flux of the source in the V-band is used, for *Swift*-BAT the X-ray flux, for 2MRS the IR flux, and for G10 the K-band flux. The fluxes are thus used as weights, so that sources with higher flux contribute more UHECRs. This approach is very similar to the approach used in Abreu et al. (2010), where

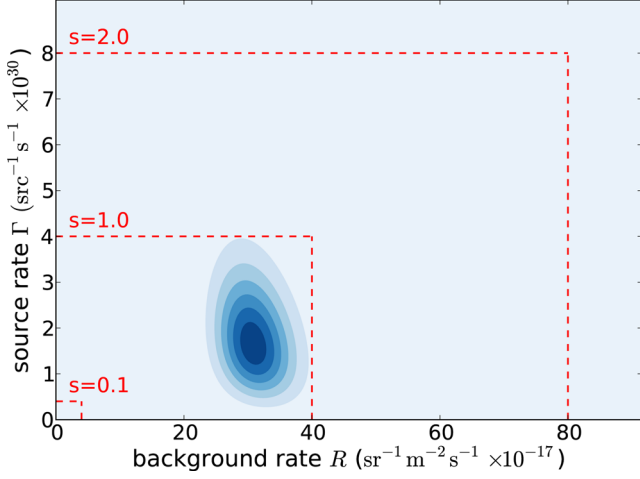


Figure 4. Example likelihood and priors for three values of s .

fluxes were used to weigh the sources from the *Swift*-BAT and 2MRS catalogues in the same way. Incorporating the fluxes into the formalism, we obtain the expression

$$\frac{dN_s(E_{\text{obs}} \geq E_{\text{thres}})}{dt dA} = \frac{\Gamma}{4\pi D_{\text{Cen}}^2} \frac{F_s}{F_{\text{Cen}}} \left[\frac{E_{\text{emit}}(E_{\text{thres}})}{E_{\text{min}}} \right]^{-\gamma}, \quad (7)$$

where D_{Cen} is the distance to Centaurus A.

4 STATISTICAL FORMALISM

Given a sample of UHECRs arrival directions, we would like to determine the fraction of these rays that have come from a set of sources under consideration. To do so, we use a two-component parametric model characterized by two rates: the source rate Γ and the isotropic background rate R . As elaborated in Section 3.4, we use the source rate of Centaurus A as the reference value of Γ . We obtain a joint posterior distribution for the two rates:

$$Pr(\Gamma, R|\mathbf{d}) = \frac{Pr(\Gamma, R) Pr(\mathbf{d}|\Gamma, R)}{\int_0^\infty \int_0^\infty Pr(\Gamma, R) Pr(\mathbf{d}|\Gamma, R) d\Gamma dR}, \quad (8)$$

where $Pr(\Gamma, R)$ is the prior distribution for Γ and R , and $Pr(\mathbf{d}|\Gamma, R)$ is the likelihood (i.e. the probability of obtaining the data set \mathbf{d} given values of Γ and R).

4.1 Prior

We adopt a uniform prior over Γ and R , with $\Gamma \geq 0$, $R \geq 0$. This plausibly encodes our ignorance of the two parameters, and, unlike maximum entropy priors, includes a possible value of 0 for both parameters. The maximum values of Γ and R are denoted as Γ_{max} and R_{max} . We have conducted our analysis for flat priors of varying width, using a variable width parameter s . The expression for the prior can be written as

$$Pr(\Gamma, R|\mathbf{d}, M_2) = \begin{cases} \frac{1}{s^2 \Gamma_{\text{max}} R_{\text{max}}} & \text{if } R < R_{\text{max}}, \Gamma < \Gamma_{\text{max}} \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

The values of Γ_{max} and R_{max} were chosen differently depending on the likelihood, so that Bayes factors for different likelihoods could be compared. Fig. 4 shows an example likelihood and shows

that Γ_{max} and R_{max} are chosen in such a way that when $s = 1$ and, the prior covers the 99.7 per cent credible region implied by the combination of the likelihood and an infinitely broad uniform prior. This gives a data driven scaling for the rates. The priors and their dependence on s are discussed in more detail in Appendix B.

4.2 The likelihood

To compute the likelihood, we use a ‘counts in cells’ approach, in which the sky is divided into $1800 \times 3600 = 6480\,000$ pixels, that are distributed uniformly in right ascension and declination. Thus, the data set \mathbf{d} can be rewritten as a set of counts in each pixel $\{N_{c,p}\}$.

The likelihood $Pr(\mathbf{d}|\Gamma, R)$ is then given by a product of the individual Poisson likelihoods in each pixel,

$$Pr(\mathbf{d}|\Gamma, R) = \prod_{p=1}^{N_p} \frac{(\bar{N}_{\text{src},p} + \bar{N}_{\text{bkg},p})^{N_{c,p}} \exp[-(\bar{N}_{\text{src},p} + \bar{N}_{\text{bkg},p})]}{N_{c,p}!}, \quad (10)$$

where $\bar{N}_{\text{src},p}$ and $\bar{N}_{\text{bkg},p}$ are the expected counts in pixel p due to sources and background, respectively. The expected number of counts in pixel p that are contributed by the background is

$$\bar{N}_{\text{bkg},p} = R \int_p \frac{d\epsilon}{d\Omega} d\Omega_{\text{obs}}, \quad (11)$$

where the integral is over the pixel p , and $d\epsilon/d\Omega$ is the relative exposure (Section 2). The expected number of source originating events in pixel p

$$\bar{N}_{\text{src},p} = \sum_{s=1}^{N_s} \frac{dN_s(E_{\text{obs}} \geq E_{\text{thres}})}{dt dA} \int_p \frac{d\epsilon}{d\Omega} Pr(\vec{r}_{\text{obs}}|\vec{r}_s) d\Omega_{\text{obs}}, \quad (12)$$

where the sum is over the sources, $Pr(\vec{r}_{\text{obs}}|\vec{r}_s)$ is the vMF distribution (equation 5), and $dN_s(E_{\text{obs}} \geq E_{\text{thres}})/dt dA$ is the observed UHECR flux discussed in Section 3.4. Inserting equations (11) and (12) into equation (10), we arrive at the full likelihood.

The positional dependence of $\bar{N}_{\text{bkg},p}$ follows the relative exposure of PAO, as shown in Fig. 1. The positional dependence of $\bar{N}_{\text{src},p}$ depends both on the PAO exposure and on the distribution of sources in the given catalogue. Fig. 2 shows the dependence for the four catalogues that are used in this study. The dependence is dominated by the distribution of local AGNs, by far the strongest source being Centaurus A ($l = 309.5^\circ$, $b = 19.4^\circ$), which previously studies (e.g. Abraham et al. 2007) have suggested as the dominant UHECR source.

The expression for the likelihood can be rearranged to reduce the total number of computations, as described in Appendix A.

4.3 The source fraction

The source fraction¹ is defined as the fraction of the UHECRs that are expected to have originated at the sources in whichever catalogue is under consideration and is given by

$$F_{\text{src}}(\Gamma, R) = \frac{\sum_{p=1}^{N_p} \bar{N}_{\text{src},p}}{\sum_{p=1}^{N_p} \bar{N}_{\text{src},p} + \bar{N}_{\text{bkg},p}}. \quad (13)$$

¹ The source fraction F_{src} is equivalent to the AGN fraction F_{AGN} used in Watson et al. (2011) but now generalized to allow for non-AGN progenitors.

The posterior for F_{src} can be calculated from the posterior over the rates as

$$Pr(F_{\text{src}}|\mathbf{d}) = \int_0^{\Gamma_{\text{max}}} \int_0^{R_{\text{max}}} Pr(\Gamma, R|\mathbf{d}) \delta_D[F_{\text{src}} - F_{\text{src}}(\Gamma, R)] d\Gamma dR. \quad (14)$$

$Pr(F_{\text{src}}|\mathbf{d})$ is insensitive to R_{max} and Γ_{max} provided they are sufficiently large.

4.4 Model comparison

We would like to compare model M_1 where all the UHECRs are drawn from a uniform distribution with model M_2 , where the UHECRs are derived from a combination of a background and a source originating component. To do this, we conduct a Bayesian model comparison. For a data set \mathbf{d} and two models, M_1 and M_2 , the ratio of the marginal likelihoods for the two models, termed the Bayes factor, is

$$B_{12} = \frac{Pr(\mathbf{d}|M_1)}{Pr(\mathbf{d}|M_2)}. \quad (15)$$

In the specific case that is considered here, the models are nested: when $\Gamma = 0$, model M_2 reduces to model M_1 . A general expression of the Bayes factor in this situation is

$$B_{12} = \frac{\int Pr(R|M_1) Pr(\mathbf{d}|R, M_1) dR}{\int Pr(\Gamma, R|M_2) Pr(\mathbf{d}|\Gamma, R, M_2) d\Gamma dR}. \quad (16)$$

It can be shown (Dickey 1971) that in the case of such nested models, the expression reduces to

$$B_{12} = \frac{Pr(\Gamma = 0|\mathbf{d}, M_2)}{Pr(\Gamma = 0|M_2)}. \quad (17)$$

This expression is known as the Savage–Dickey Density Ratio, or SDDR. Qualitatively, this expression means that the nested uniform model is preferred if, within the context of the more complicated model, the data result in an increased probability that $\Gamma = 0$.

5 SIMULATIONS

In order to investigate the constraining power of a data set of 69 events, we apply the method to simulated data sets. We use two extreme cases as follows.

(i) Uniform arrival directions: these rays were drawn from a probability distribution that followed the PAO exposure.

(ii) UHECRs originating at sources from a catalogue: we conducted simulations for all four of the catalogues. In each catalogue, the sources were weighted by their fluxes and the PAO exposure. Random sources were then selected, and the propagation model of Section 3 was used to propagate rays from the sources to the Earth.

The posteriors for the source and background rates, as well as the posteriors for the source fraction, are summarized in Fig. 5. The posteriors for the uniform and source centred cases are completely disjoint, which demonstrates that in extreme scenarios, where all UHECRs originate either from a uniform background or from a source catalogue, a data set of 69 events should be sufficient to distinguish between the two models. Fig. 5 also shows the Bayes factors as functions of s for the two cases. The Bayes factors B_{21} that are displayed are the inverses of the SDDR given in equation (17), and favour the more complex model for Bayes factors > 1 .

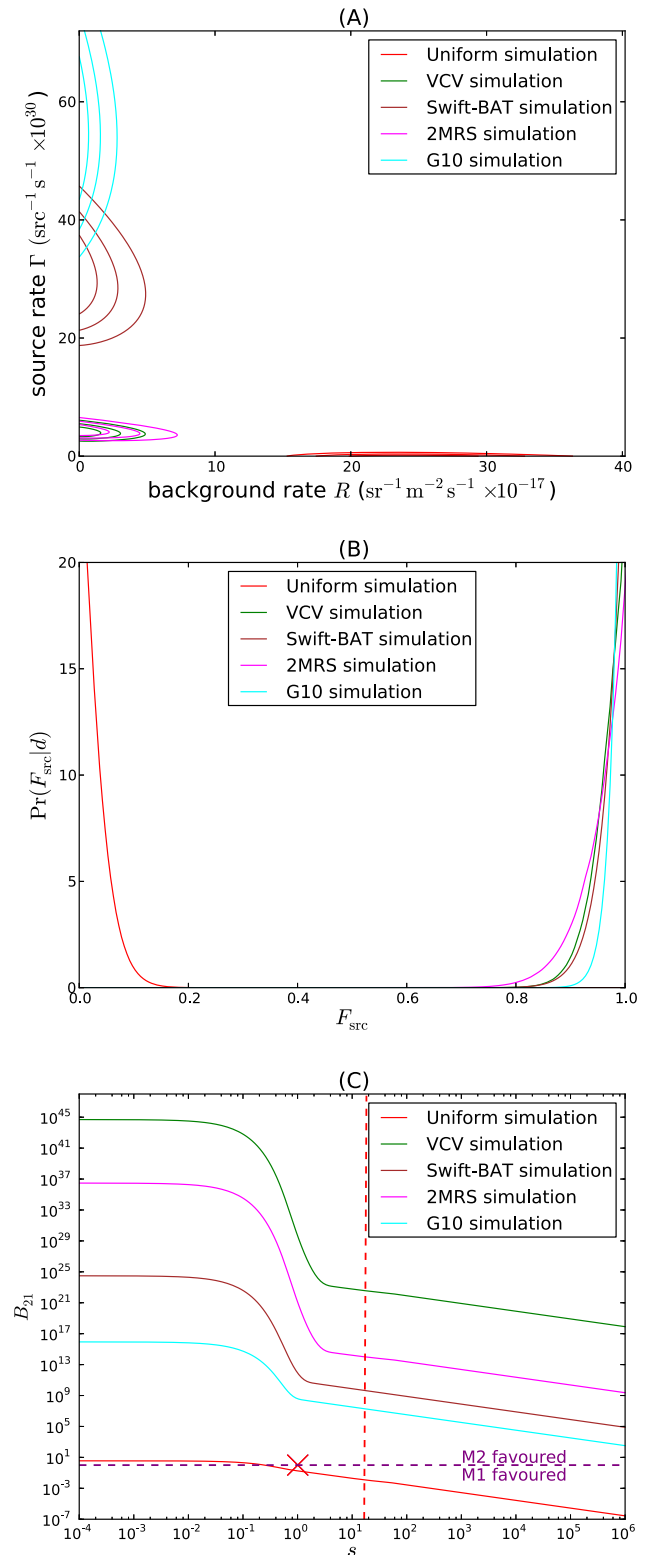


Figure 5. Results from simulations: uniform UHECRs, and UHECRs originating at sources from the four catalogues. In all cases, 69 events are used. (A) Posteriors for Γ and R . The contours are the 68.3, 95.4 and 99.7 per cent highest posterior density credible regions. (B) Posteriors for the source fraction. (C) Plot of Bayes factors, B_{21} , as a function of the hyperparameter s . In (C), the \times -mark and the vertical line signify the minimum and the maximum values of the physically plausible range of s . The minimum and maximum values that are displayed correspond to the uniform simulation.

To assess the results of the Bayes factor simulations, we can derive a rough range of plausible values of s from physical models, and then look at the behaviour of the Bayes factors at those physically plausible values. Plausible models of UHECR injection predict that the UHECR luminosity of a source like Centaurus A is of the order of $2.9 \times 10^{39} \text{ erg s}^{-1} \simeq 1.81 \times 10^{51} \text{ eV s}^{-1}$ (Fraija et al. 2012). If this is taken as the typical UHECR luminosity of a source, then for a UHECR energy range of $(5.7\text{--}100) \times 10^{19} \text{ eV}$, the range of source rates can be calculated by dividing the UHECR luminosity by the limiting values of this range. The result of this calculation is a range of source rates Γ of roughly $(2\text{--}33) \times 10^{30} \text{ s}^{-1}$. The values of s corresponding to this range have been marked on Fig. 5. (The values are slightly different for each of the simulations. For the sake of clarity, only the values for the uniform simulation are displayed, the others being broadly similar.) For the sourced case, model M_2 is strongly favoured for all physically plausible values of s , while for the uniform case, the simple uniform model M_1 is favoured for the physically plausible values.

6 RESULTS

The results of the application of the statistical methods described in Section 4 to the data described in Section 2 are shown in Figs 6 and 7. Fig. 6 contrasts the results from our analysis with the equivalent results from Watson et al. (2011), and with the results for an intermediate case. The use of a more refined propagation model leads to a higher posterior probability for lower source rates. The reason for that is that in Watson’s propagation model, the energy loss length is constant and very small (Fig. 3). UHECRs experience more drastic energy loss than in the more realistic model, which leads to more distant AGNs being excluded as plausible source candidates. As fewer sources are included, a higher source rate is required to generate the same sample of UHECRs.

The inclusion of 69 events reduces the extent to which the non-uniform model is favoured. This is evident from the posterior of the source fraction, and also from the behaviour of B_{21} . This result agrees with the results of Abreu et al. (2010), which reported that the full 69 events yield lower evidence of anisotropy than the earlier study Abraham et al. (2007), which analysed 27 events.

Fig. 7 shows results for all four of the source catalogues, and for all values of the smearing parameter. Displayed are the posteriors for the source fraction, as well as plots of B_{21} against s . The constraints on the source fraction for all cases are shown in Table 1. The figures and table show that for greater smearing, the range of plausible values of F_{src} is increased, and the most probable value of the source fraction is higher than for the fiducial model of $\sigma = 3^\circ$. The reason is that for greater magnetic deflection, the UHECR intensity distribution becomes more uniform, so that the uniform and mixed models become more difficult to distinguish, and a greater range of F_{src} values become viable.

The plots of B_{21} demonstrate that for all physically plausible prior ranges of the model parameters, the fully isotropic model is disfavoured. The form of the dependence of B_{21} on s is elaborated upon in Appendix B.

These results for the VCV, *Swift*-BAT, and 2MRS catalogues can be compared with the results of Abreu et al. (2010), who used a correlation-based analysis on the VCV catalogue that mirrored the analysis in Abraham et al. (2007). Abreu et al. (2010) reported a correlation of (38_{-6}^{+7}) per cent between UHECRs and sources from the VCV catalogue, which was considerably lower than the (69_{-13}^{+11}) per cent correlation that was reported in Abraham et al. (2007). This reduction in the correlation is consistent with our find-

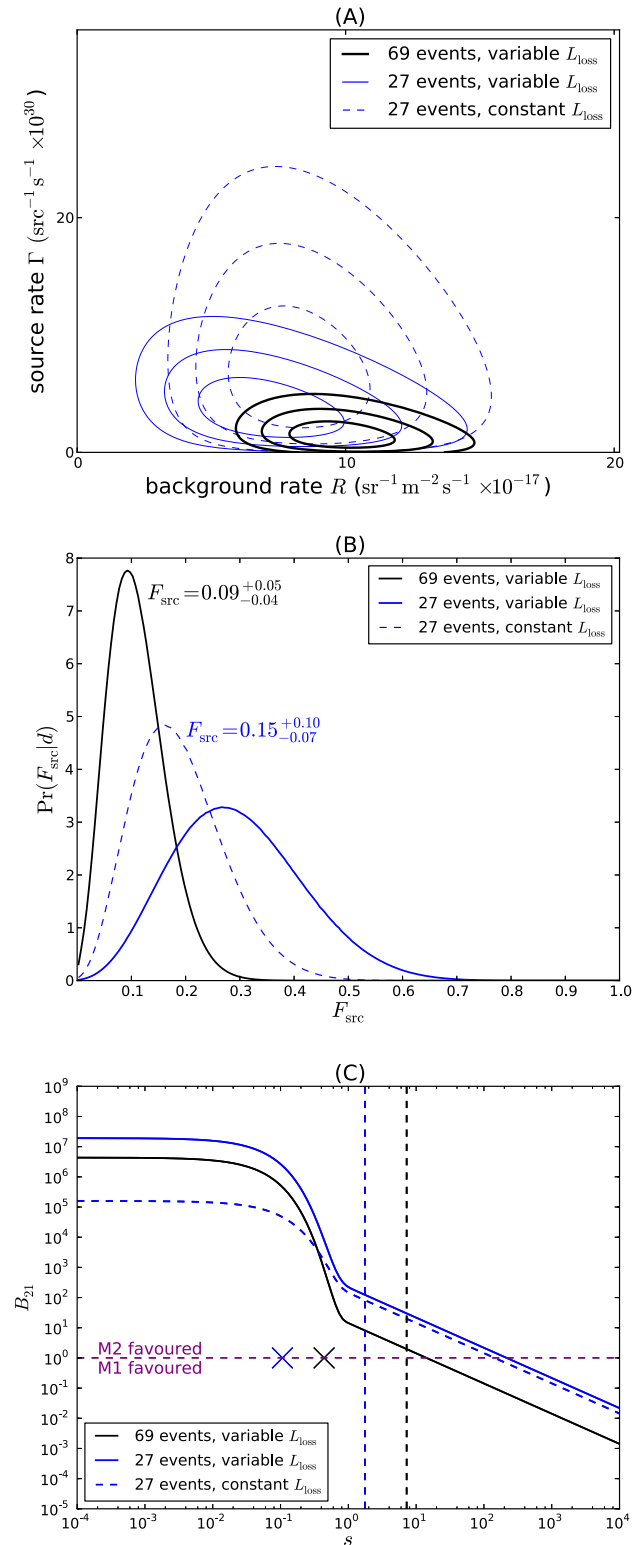


Figure 6. Results for $\sigma = 3^\circ$, and the sources from the VCV catalogue. Results for 27 and 69 events, and for constant and variable loss lengths are displayed. (A) Posteriors for the source and background rates. The contours are the 68.3, 95.4 and 99.7 per cent, the highest posterior density credible regions. (B) Posterior for the source fraction. (C) Plot of Bayes factors B_{21} as a function of the hyperparameter s . In (C), physically plausible ranges of s are shown for the cases of 27 events (blue) and 69 events (black), with a variable loss length. The \times -marks and the vertical lines signify the minimum and the maximum values of the physically plausible ranges of s .

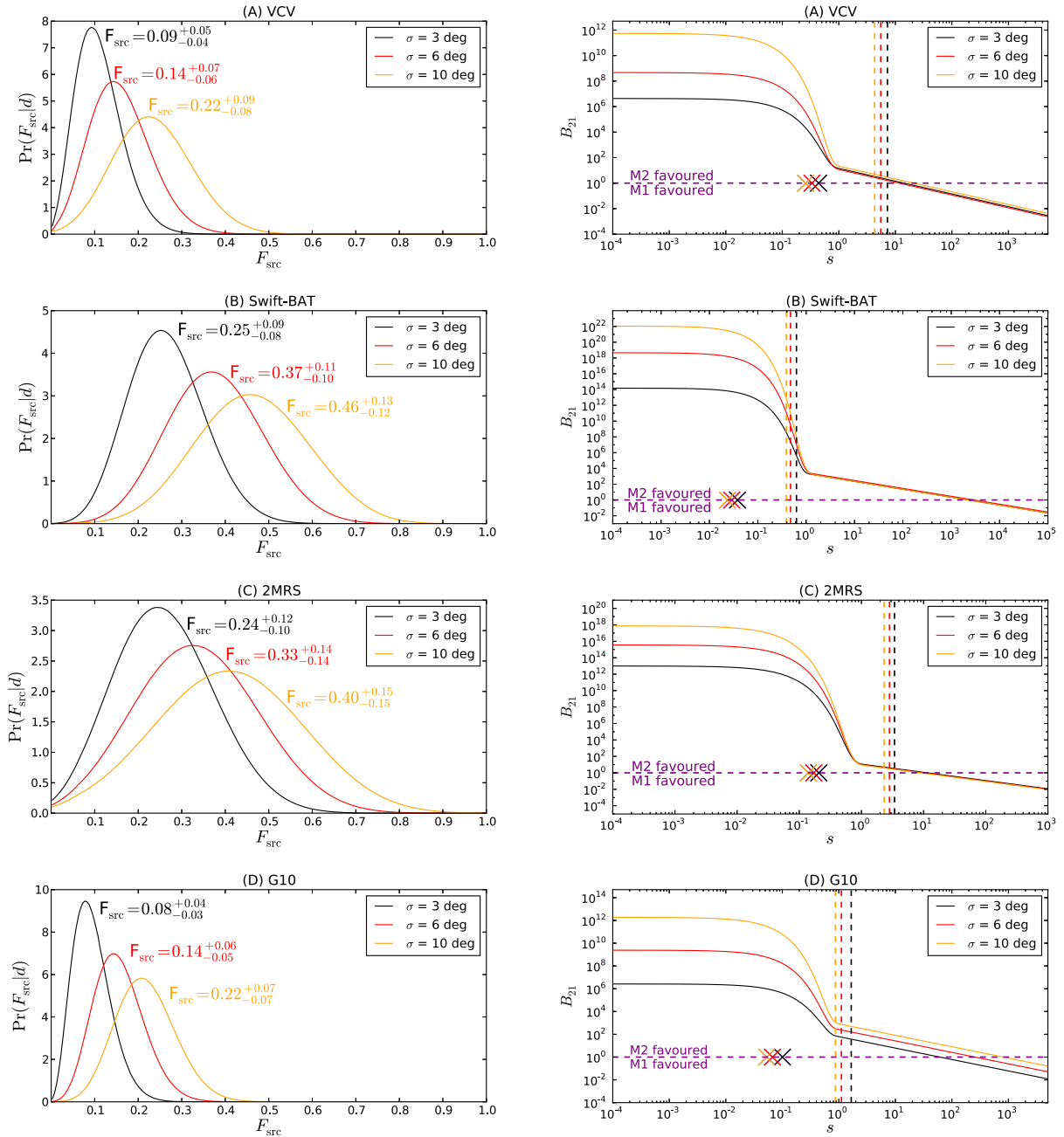


Figure 7. Posteriors of the source fraction, and plots of B_{21} against the hyperparameter s , for the three smearing angles $\sigma = 3^\circ$, 6° and 10° , and for the three source catalogues (A) VCV, (B) *Swift*-BAT, and (C) 2MRS. The plots of B_{21} show physically plausible ranges of s : The \times -marks and the vertical lines signify the minimum and the maximum values of these ranges.

Table 1. Maximum a posteriori estimates and 68 per cent credible intervals for F_{src} .

Catalogue	$\sigma = 3^\circ$	$\sigma = 6^\circ$	$\sigma = 10^\circ$
VCV	$0.09^{+0.05}_{-0.04}$	$0.14^{+0.07}_{-0.06}$	$0.22^{+0.09}_{-0.08}$
<i>Swift</i> -BAT	$0.25^{+0.09}_{-0.08}$	$0.37^{+0.11}_{-0.10}$	$0.46^{+0.13}_{-0.12}$
2MRS	$0.24^{+0.12}_{-0.10}$	$0.33^{+0.14}_{-0.14}$	$0.40^{+0.15}_{-0.15}$
G10	$0.08^{+0.04}_{-0.03}$	$0.14^{+0.06}_{-0.05}$	$0.22^{+0.07}_{-0.07}$

ings that the source fraction is reduced as we increase the data set from 27 to 69 events. In addition to these correlation based methods, Abreu et al. (2010) conducted a likelihood-based study similar to the analysis presented here, where the likelihood was taken as a probability map of arrival directions of UHECRs, parametrized by a magnetic smoothing angle σ and a fraction of isotropic rays f_{iso} , which is equivalent to $1 - F_{\text{src}}$. These likelihood-based studies were conducted for the *Swift*-BAT and 2MRS catalogues. For the 2MRS case, the maximum likelihood values of f_{iso} and σ are reported as 0.56° and 7.8° , respectively. The σ value lies between our chosen smearing angles 6° and 10° . The value for f_{iso} corresponds

to a value of F_{src} of 0.44, which is consistent with our F_{src} credible intervals for these chosen smearing angles. For the case of *Swift*-BAT, the maximum likelihood value of f_{iso} is given as 0.64, which corresponds to a source fraction of 0.36. The maximum likelihood estimate of the smearing angle is reported as 1.5° , which is lower than our minimum chosen value of 3° . Despite the difference between the angles, a F_{src} value of 0.36 can still be considered broadly consistent with the 68 per cent credible interval for 3° , $0.25^{+0.09}_{-0.08}$.

Our results for the G10 catalogue can be compared with the work of Soiaporn et al. (2013). That analysis involved the full data set of 69 events, and found evidence for small but non-zero values of F_{src} , of the order of a few per cent to 20 per cent. This is broadly consistent with our results, which suggest that values of $F_{\text{src}} \lesssim 0.3$ are the most probable for all values of the smearing parameter.

These results assume a pure proton composition of UHECRs, although there is some evidence that this is not the case at the highest energies. HiRes and Telescope Array claim that the composition is consistent with being proton-dominated for the entire energy range (Abbasi et al. 2010b; Jui et al. 2012). Evidence from PAO, on the other hand, suggests a gradual transition from light protonic composition to heavier elements such as iron for $E_{\text{arr}} \gtrsim 4 \times 10^{18}$ eV (Abraham et al. 2010). Significantly, greater magnetic smearing is expected for heavier nuclei, as the angular scale of the deflection is proportional to the atomic number of the UHECR (Harari et al. 2002). The energy losses depend on the composition of the nucleus (Stanev 2009) and for most heavier nuclei the minimum values of the loss length are significantly lower than for protons, at least for the highest energies ($E \gtrsim 10^{20.5}$ eV).

The net effect of heavier composition on the results is difficult to assess a priori. The increased magnetic smearing on its own should decrease the extent to which the non-uniform model is favoured in the analysis. However, as the energy loss is greater, for at least some of the energy range, it could lead to a reduced GZK horizon, within which the source distribution would be more inhomogeneous, thus, increasing the extent to which the non-uniform model is favoured. Closer sources would also reduce the angular scale over which UHECRs are smeared, which could compensate for the increased magnetic deflection. The way to resolve this a priori uncertainty would be to modify our model to take into account heavier nuclei and to apply the model to the data, or to mock data sets. Our propagation model cannot easily be modified to account for the fact that the composition of the heavier nuclei changes through the scattering processes, although a different approach would be to incorporate a computational framework such as GALPROP (Moskalenko & Seo 2014) into our statistical formalism, which could be explored in future research.

7 CONCLUSIONS

We have performed a Bayesian analysis of the 69 UHECRs detected by the PAO with energies $E_{\text{obs}} > 5.7 \times 10^{19}$ eV to determine the fraction of these UHECRs that originated from catalogues of plausible UHECR sources. The sources considered were AGNs from the VCV, *Swift*-BAT, and G10 catalogues, and galaxies from the 2MRS catalogue.

For the fiducial magnetic smearing parameter of $\sigma = 3^\circ$, we report 68 per cent credible intervals for the source fraction of $0.09^{+0.05}_{-0.04}$, $0.25^{+0.09}_{-0.08}$, $0.08^{+0.04}_{-0.03}$, and $0.24^{+0.12}_{-0.10}$ for the VCV, *Swift*-BAT, G10 and 2MRS catalogues, respectively. For all physically plausible values of the model parameters, the fully uniform model is disfavoured. The results of our study are in broad agreement with previous work on this subject, such as Watson et al. (2011), Abreu et al. (2010), and

Soiaporn et al. (2013). The credible intervals for the VCV catalogue are lower than the analogous credible intervals from Watson et al. (2011), which used a similar method to analyse 27 PAO events. This is consistent with earlier studies: Abreu et al. (2010), which analysed 69 events, reported a lower signal of anisotropy than the earlier study Abraham et al. (2007), which used 27 events.

We will extend this Bayesian framework to include the arrival energies of the UHECRs as well as the arrival directions.

It is expected that future experiments will produce data sets that will be sufficiently large for our Bayesian method (and other statistical approaches; see e.g. Rouillé d'Orfeuil et al. 2014) to detect even the weak clustering expected if the UHECRs have come from nearby sources. PAO is continuing to take data and is expected to produce a sample of ~ 250 UHECRs over its first decade of operations. Looking further ahead, the planned Japanese Experiment Module Extreme Universe Space Observatory (JEM-EUSO; Adams et al. 2013) on the International Space Station (ISS) is scheduled for launch in 2017 and is expected to detect ~ 200 UHECRs annually over its five year lifetime.

ACKNOWLEDGEMENTS

We thank an anonymous referee for making suggestions which allowed us to improve this paper. AK was supported by a Science & Technology Facilities Council studentship.

REFERENCES

- Aab A. et al., 2015, *ApJ*, 804, 15
 Abbasi R. U. et al., 2008a, *Astropart. Phys.*, 30, 175
 Abbasi R. U. et al., 2008b, *Phys. Rev. Lett.*, 100, 101101
 Abbasi R. et al., 2010a, *ApJ*, 713, L64
 Abbasi R. U. et al., 2010b, *Phys. Rev. Lett.*, 104, 199902
 Abraham J. et al., 2007, *Science*, 318, 938
 Abraham J. et al., 2008a, *Astropart. Phys.*, 29, 188
 Abraham J. et al., 2008b, *Phys. Rev. Lett.*, 101, 061101
 Abraham J. et al., 2009, preprint (arXiv:0906.2347)
 Abraham J. et al., 2010, *Phys. Lett. B*, 685, 239
 Abreu P. et al., 2010, *Astropart. Phys.*, 34, 314
 Abu-Zayyad T. et al., 2012, *ApJ*, 757, 26
 Adams J. H., Jr, et al., 2013, preprint (arXiv:1307.7071)
 Baumgartner W. H. et al., 2010, *ApJS*, submitted
 De Domenico M., Insolia A., 2013, *J. Phys. G: Nucl. Phys.*, 40, 015201
 Dickey J., 1971, *Ann. Math. Stat.*, 42, 204
 Dolag K., Grasso D., Springel V., Tkachev I., 2005, *J. Cosmol. Astropart. Phys.*, 1, 9
 Farrar G. R., Jansson R., Feain I. J., Gaensler B. M., 2013, *J. Cosmol. Astropart. Phys.*, 1, 023
 Fraija N. et al., 2012, *ApJ*, 753, 40
 George M. R., Fabian A. C., Baumgartner W. H., Mushotzky R. F., Tueller J., 2008, *MNRAS*, 388, L59
 Goulding A. D., Alexander D. M., Lehmer B. D., Mullaney J. R., 2010, *MNRAS*, 406, 597 (G10)
 Gregory P. C., 2010, *Bayesian Logical Data Analysis for the Physical Sciences*. Cambridge Univ. Press, Cambridge, p. 376
 Greisen K., 1966, *Phys. Rev. Lett.*, 16, 748
 Harari D., Mollerach S., Roulet E., Sánchez F., 2002, *J. High Energy Phys.*, 3, 045
 Hillas A. M., 1968, *Can. J. Phys.*, 46, 623
 Huchra J. P. et al., 2012, *ApJ*, 199, 26
 Ivanov A. A., 2009, *Nuc. Phys. B*, 190, 204
 Jui C. C. H. et al., 2012, *J. Phys. Conf. Ser.*, 404, 012037
 Kalmykov N. N., Khrenov B. A., Kulikov G. V., Zotov M. Y., 2013, *J. Phys. Conf. Ser.*, 409, 012100
 Keivani A., Farrar G. R., Sutherland M., 2015, *Astropart. Phys.*, 61, 47

Kotera K., Olinto A. V., 2011, *Ann. Rev. Astron. Astrophys.*, 49, 119
 Letessier-Selvon A., Stanev T., 2011, *Rev. Mod. Phys.*, 83, 907
 Letessier-Selvon A. et al., 2014, *Braz. J. Phys.*, 44, 560
 Medina Tanco G. A., de Gouveia Dal Pino E. M., Horvath J. E., 1998, *ApJ*, 492, 200
 Moskalenko I. V., Seo E., 2014, *Adv. Space Res.*, 53, 1377
 Nagar N. M., Matulich J., 2010, *A&A*, 523, A49
 Pe'er A., Murase K., Mészáros P., 2009, *Phys. Rev. D*, 80, 123018
 Rouillé d'Orfeuille B., Allard D., Lachaud C., Parizot E., Blaksley C., Nagataki S., 2014, *A&A*, 567, A81

Sigl G., Miniati F., Enßlin T. A., 2004, *Phys. Rev. D*, 70, 043007
 Soiaporn K., Chernoff D., Loredó T., Ruppert D., Wasserman I., 2013, *Ann. Appl. Stat.*, 7, 1249
 Stanev T., 2009, *New J. Phys.*, 11, 13
 Tinyakov P. G., Tkachev I. I., 2001, *Sov. J. Exp. Theor. Phys. Lett.*, 74, 445
 Véron-Cetty M.-P., Véron P., 2006, *A&A*, 455, 773
 Watson L. J., Mortlock D. J., Jaffe A. H., 2011, *MNRAS*, 418, 206
 Zatsepin G., Kuzmin V., 1966, *J. Exp. Theor. Phys. Lett.*, 4, 78

APPENDIX A: NUMERICAL EVALUATION OF THE LIKELIHOOD

The likelihood, as given in equation (10), is a product over Poisson likelihoods for the individual pixels,

$$Pr(\mathbf{d}|\Gamma, R) = \prod_{p=1}^{N_p} \frac{(\bar{N}_{\text{src},p} + \bar{N}_{\text{bkg},p})^{N_{c,p}} \exp[-(\bar{N}_{\text{src},p} + \bar{N}_{\text{bkg},p})]}{N_{c,p}!}, \quad (\text{A1})$$

where the product is over the pixels, $N_{c,p}$ is the number of counts in pixel p , and $\bar{N}_{\text{bkg},p}$ and $\bar{N}_{\text{src},p}$ are the expected numbers of counts from the background and sources in pixel p . This expression for the likelihood proved to be inefficient for use, as it required a great number of computations: The total number of pixels was $N_p = 1800 \times 3600 = 6480\,000$. If a $\Gamma \times R$ grid of 100×100 is used, a total of 64 800 000 000 calculations would be required.

The total number of calculations can be greatly reduced by rearranging the expression. For a given data set, we can separate the product of equation (A1) into a product over those pixels that include an event, $\{q\}$, and pixels that do not, $\{r\}$. Using the fact that $N_q = 1$ for all $\{q\}$ and $N_r = 0$ for all $\{r\}$, we can write

$$Pr(\mathbf{d}|\Gamma, R) = \prod_{r=1}^{N_r} \exp[-(\bar{N}_{\text{src},r} + \bar{N}_{\text{bkg},r})] \times \prod_{q=1}^{N_q} (\bar{N}_{\text{src},q} + \bar{N}_{\text{bkg},q}) \exp[-(\bar{N}_{\text{src},q} + \bar{N}_{\text{bkg},q})] \quad (\text{A2})$$

$$= \exp[-(\Gamma \Sigma_{\text{src}} + R \Sigma_{\text{bkg}})] \times \prod_{q=1}^{N_q} (\bar{N}_{\text{src},q} + \bar{N}_{\text{bkg},q}) \exp[-(\bar{N}_{\text{src},q} + \bar{N}_{\text{bkg},q})]. \quad (\text{A3})$$

where $\Sigma_{\text{src}} = \sum_{r=1}^{N_r} m_{\text{src},r}$ and $\Sigma_{\text{bkg}} = \sum_{r=1}^{N_r} m_{\text{bkg},r}$, and $m_{\text{src},p}$ and $m_{\text{bkg},p}$ are two pixelized maps obeying the equations

$$\bar{N}_{\text{src},p} = \Gamma m_{\text{src},p} \quad (\text{A4})$$

$$\bar{N}_{\text{bkg},p} = R m_{\text{bkg},p}. \quad (\text{A5})$$

Thus, the initial expression has been rearranged in such a way that the vast majority of Poisson calculations is contained within the sums Σ_{src} and Σ_{bkg} . These sums can be calculated in advance for the entire grid of Γ and R . This greatly reduces the total number of calculations required for equation (A1), and speeds up the full calculation by a factor of $\sim 10^5$.

APPENDIX B: MODEL COMPARISON AND PRIOR SENSITIVITY

The Bayes factor that was discussed in Section 4.4 is comparing two models: A simple model M_1 of uniform UHECRs, and a more complex model M_2 that has both uniform and sourced UHECRs. As explained in the section, due to M_1 being nested within M_2 , the expression for the Bayes factor reduces to

$$B_{12} = \frac{\int Pr(\Gamma = 0, R|\mathbf{d}, M_2) dR}{\int Pr(\Gamma = 0, R|M_2) dR} = \frac{Pr(\Gamma = 0|\mathbf{d}, M_2)}{Pr(\Gamma = 0|M_2)}, \quad (\text{B1})$$

where Γ and R are the background and source rates and \mathbf{d} are the data. Qualitatively, the expression means that the nested uniform model is preferred if, within the context of the more complex model, the data result in an increased probability that $\Gamma = 0$. A uniform prior was used, given by

$$Pr(\Gamma, R|\mathbf{d}, M_2) = \begin{cases} \frac{1}{s^2 \Gamma_{\text{max}} R_{\text{max}}} & \text{if } R < R_{\text{max}}, \Gamma < \Gamma_{\text{max}} \\ 0 & \text{otherwise} \end{cases} \quad (\text{B2})$$

where s is the hyperparameter that determines the width of the prior. As illustrated in Fig. B1, Γ_{max} and R_{max} have been chosen in such a way that when $s = 1$, the prior covers the 99.7 per cent credible region implied by the combination of the likelihood and an infinitely broad uniform prior. To explain the dependence of the Bayes factor on s , three illustrative cases are used: The case of a simple Gaussian likelihood, the case of the Poisson product likelihood of equation (10), and the likelihood of On/Off measurements.

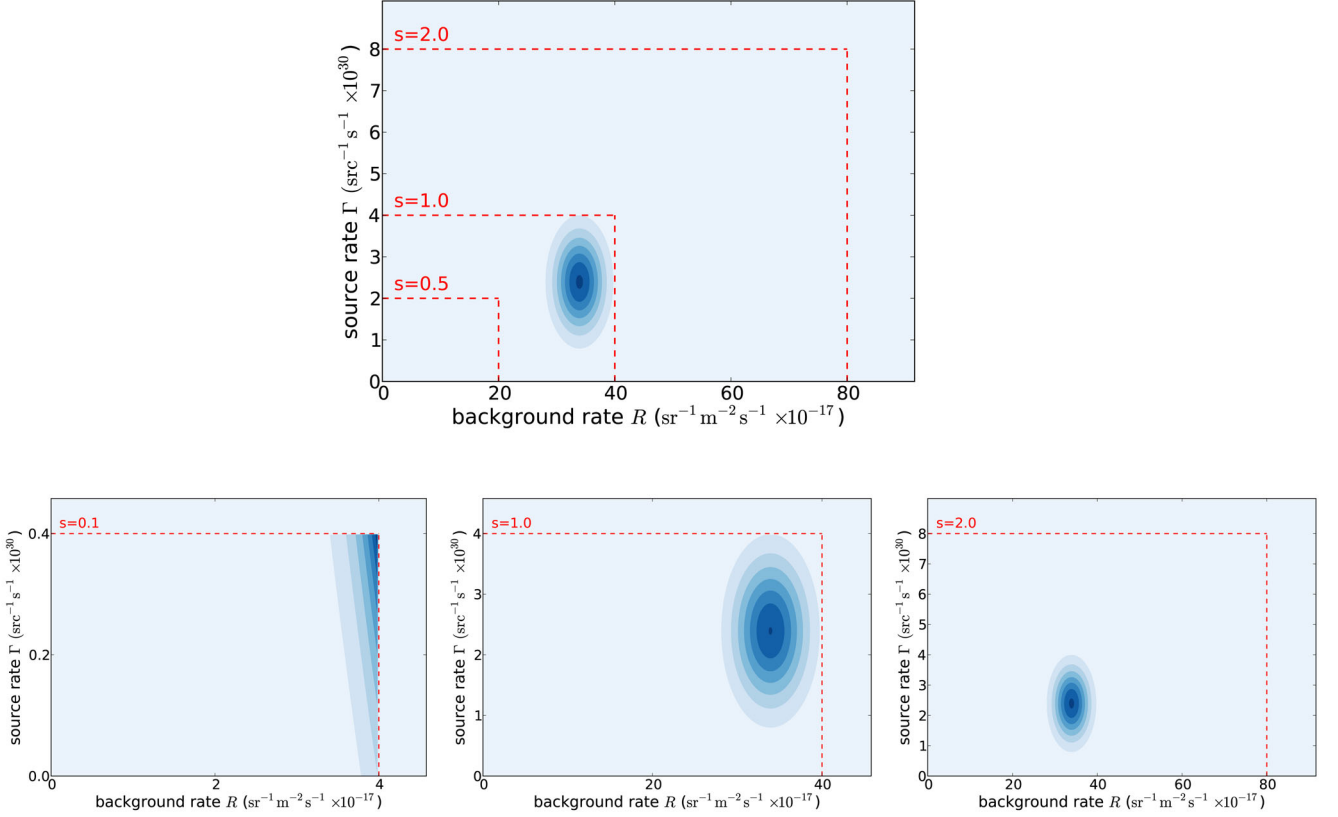


Figure B1. Upper panel: Example Gaussian likelihood. The red lines denote prior regions for three different values of the hyperparameter s . Lower panel: Posteriors for the same s values are displayed.

B1 Gaussian likelihood

We consider the case of a Gaussian likelihood given by

$$Pr(\mathbf{d}|\Gamma, R) = \frac{1}{2\pi\sigma_\Gamma\sigma_R} \exp\left[-\frac{(\Gamma - \Gamma_\mu)^2}{2\sigma_\Gamma^2}\right] \left[-\frac{(R - R_\mu)^2}{2\sigma_R^2}\right], \quad (\text{B3})$$

where Γ_μ and R_μ are the coordinates of the likelihood mean, σ_Γ and σ_R are the standard deviations on the two parameters.

This likelihood is shown in the upper panel of Fig. B1, focusing on three regions $s = 0.1, 1, 2$. These regions correspond to the regions over which the flat prior is taken for these values of the hyperparameter. The lower panel shows the posteriors $Pr(\Gamma, R|\mathbf{d}, M_2)$ for the same s values. As the priors are flat, the posteriors are equivalent to the likelihood in the prior region, normalized over the prior region.

These posteriors can be used to illustrate the dependence of the Bayes factor in equation (B1) on the hyperparameter s . For $s > 1$, the numerator $Pr(\Gamma = 0|\mathbf{d}, M_2)$ is constant, as $Pr(\Gamma, R|\mathbf{d}, M_2)$ corresponds to the normalized likelihood, and does not vary as s is increased beyond $s = 1$. The denominator $Pr(\Gamma = 0|M_2)$ falls linearly with s . Thus, we expect that for $s > 1$, B_{12} increases linearly with s .

For lower values of s , the behaviour of B_{12} is more complicated, as can be seen in the left-hand lower panel of Fig. B1. For low values of s , the likelihood becomes

$$Pr(\Gamma, R|\mathbf{d}) = \frac{1}{2\pi\sigma_\Gamma\sigma_R} e^{-\frac{\Gamma_\mu^2}{2\sigma_\Gamma^2}} e^{-\frac{R_\mu^2}{2\sigma_R^2}} \left(1 + \frac{\Gamma_\mu\Gamma}{2\sigma_\Gamma^2} + \frac{R_\mu R}{2\sigma_R^2}\right). \quad (\text{B4})$$

This means that the posterior becomes linear and increasingly flat as $s \rightarrow 0$. As the function becomes increasingly flat, the ratio in equation (B1) becomes a ratio of two normalized flat functions, so that qualitatively, we can expect it to approach unity. This can also be shown more rigorously, as for low values of Γ and R , equation (B1) reduces to

$$B_{12} = 1 - s \frac{\Gamma_\mu \Gamma_{\max}}{2\sigma_\Gamma^2}. \quad (\text{B5})$$

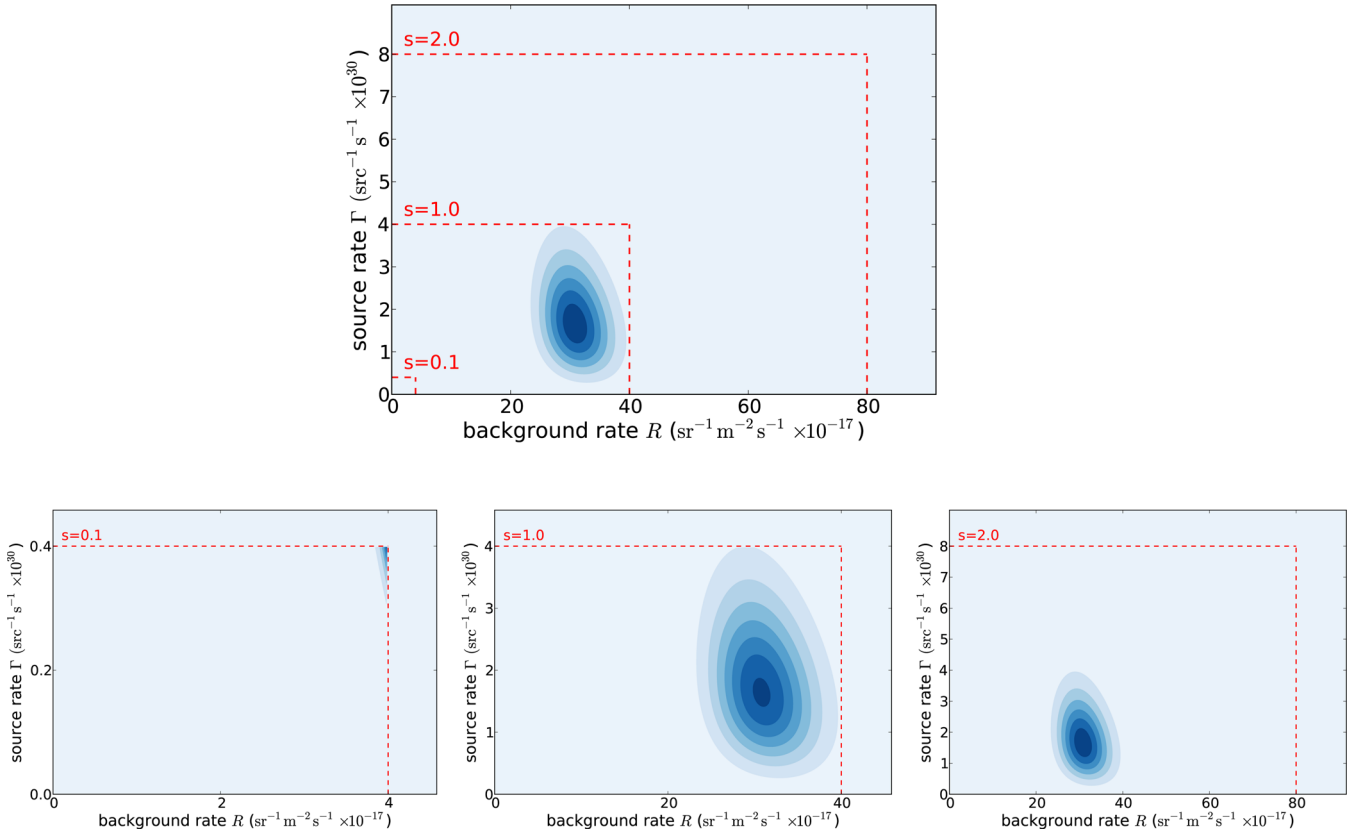


Figure B2. Upper panel: example Poisson product likelihood. The red lines denote prior regions for three different values of the hyperparameter s . Lower panel: posteriors for the same s values are displayed.

B2 Poisson product likelihood

We consider the same likelihood that was used in equation (10). The total likelihood is a product of individual Poisson likelihoods for 6 480 000 pixels, and can be written as

$$Pr(\mathbf{d}|\Gamma, R) = \prod_{p=1}^{N_p} \frac{(\bar{N}_{\text{src},p} + \bar{N}_{\text{bkg},p})^{N_{c,p}} \exp[-(\bar{N}_{\text{src},p} + \bar{N}_{\text{bkg},p})]}{N_{c,p}!}, \quad (\text{B6})$$

where $\bar{N}_{\text{src},p}$ and $\bar{N}_{\text{bkg},p}$ are the expected numbers of counts in pixel p due to source and background rates, respectively. Fig. B2 shows the likelihood, as well as three prior regions, and the posteriors calculated for these three regions.

For high values of s , the posterior looks very much like the Gaussian, so that we expect the same behaviour for the Bayes factor, including the linear behaviour for $s > 1$. A difference arises at small values of s . Here, we see that most of the posterior is concentrated at the highest values of Γ and R . As the product of equation (B6), for low values of Γ and R , reduces to

$$Pr(\mathbf{d}|\Gamma, R) = \prod_{q=1}^{N_q} (\Gamma m_{\text{src},q} + R m_{\text{bkg},q}), \quad (\text{B7})$$

where N_q is the total number of PAO events, and $m_{\text{src},r}$ and $m_{\text{bkg},r}$ are the pixelized maps that were discussed in Appendix A. As $N_q = 69$, the function becomes extremely steep in Γ and R , as Fig. B2 shows. For such a posterior, B_{12} tends to zero as $Pr(\Gamma = 0|\mathbf{d}, M_2) \ll 1$.

B3 On/Off likelihood

An additional case that is of interest in this analysis is that of On/Off measurements. In high-energy astrophysics, when a measurement is taken of the number of counts coming from a source of interest, often an auxiliary measurement is made by pointing the detector off-source. These are called the On and Off measurements, respectively. The counts that are detected in the Off measurement are thereby produced solely by the background rate R , while the counts in the On measurement are produced by both the background and the source rates Γ and R . From these two measurements, the source rate can then be estimated (e.g. Gregory 2010).

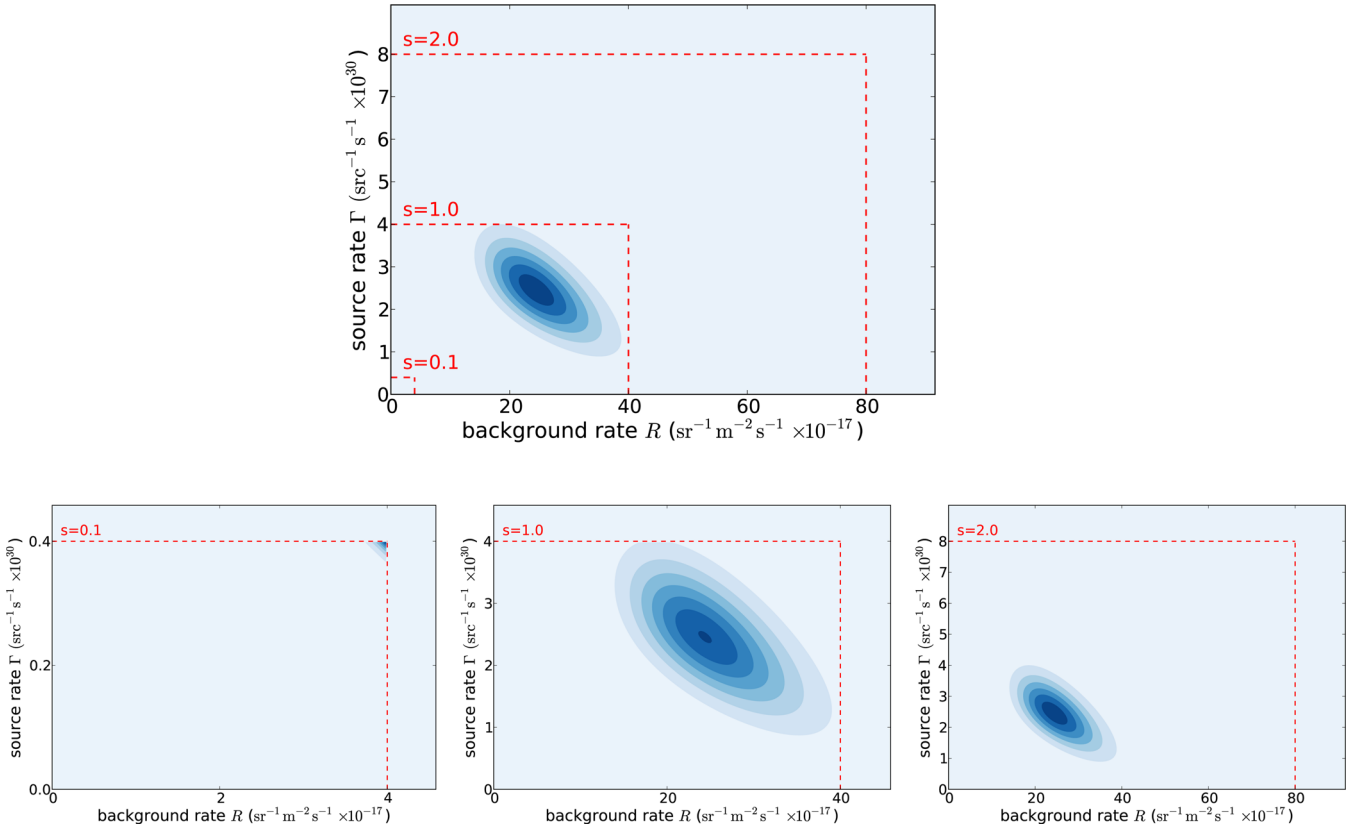


Figure B3. Upper panel: example On/Off likelihood. The red lines denote prior regions for three different values of the hyperparameter s . Lower panel: posteriors for the same s values are displayed.

The likelihood for these kinds of measurements is the product of the Poisson likelihoods of the On and Off measurements:

$$Pr(N_{\text{on}}, N_{\text{off}} | \Gamma, R) = \frac{(RT_{\text{off}})^{N_{\text{off}}} \exp(-RT_{\text{off}})}{N_{\text{off}}!} \times \frac{[(\Gamma + R)T_{\text{on}}]^{N_{\text{on}}} \exp[-(\Gamma + R)T_{\text{on}}]}{N_{\text{on}}!}, \quad (\text{B8})$$

where N_{on} and N_{off} are the numbers of counts on and off source, and T_{on} and T_{off} times the detector spends on and off the source. An example of such a likelihood is displayed in Fig. B3. The On/Off likelihood is very similar to the Poisson product likelihood, as the former can be regarded as a special case of the latter. Thus, the dependence of the Bayes factor on s can be expected to be similar to the dependence for the Poisson product case.

For the On/Off case, a standard expression for the Bayes factor has been derived (Gregory 2010), and can be written as

$$B_{21} = \frac{N_{\text{on}}!}{\Gamma_{\text{max}} T_{\text{on}} \gamma[(N_{\text{on}} + N_{\text{off}} + 1), R_{\text{max}}(T_{\text{on}} + T_{\text{off}})]} \sum_{i=0}^{N_{\text{on}}} \frac{\gamma[(N_{\text{on}} + N_{\text{off}} - i + 1), R_{\text{max}}(T_{\text{on}} + T_{\text{off}})]}{i!(N_{\text{on}} - i)!} \gamma(i + 1, \Gamma_{\text{max}} T_{\text{on}}) \left(1 + \frac{T_{\text{off}}}{T_{\text{on}}}\right)^i, \quad (\text{B9})$$

where $\gamma(s, x)$ is the lower incomplete gamma function, defined here as

$$\gamma(s, x) = \int_0^x t^{s-1} e^{-t} dt. \quad (\text{B10})$$

The standard expression reproduces the same dependence that one obtains by calculating the ratio in equation (B1). Note that this expression is for B_{21} rather than B_{12} .

Fig. B4 shows the dependence of the Bayes factor on s for the three cases. The Bayes factors that are shown in the Figure are the Bayes factors favouring the complex model, $B_{21} = 1/B_{12}$. For all three cases, B_{21} falls linearly for $s > 1$. For lower values of s , the Bayes factor for the Gaussian case approaches 1, while for the PAO and On/Off cases B_{12} becomes $\gg 1$, as the uniform model is extremely disfavoured. The Bayes factors for the On/Off case behave very similarly to the Poisson product case, as the former can be regarded as a special case of the latter.

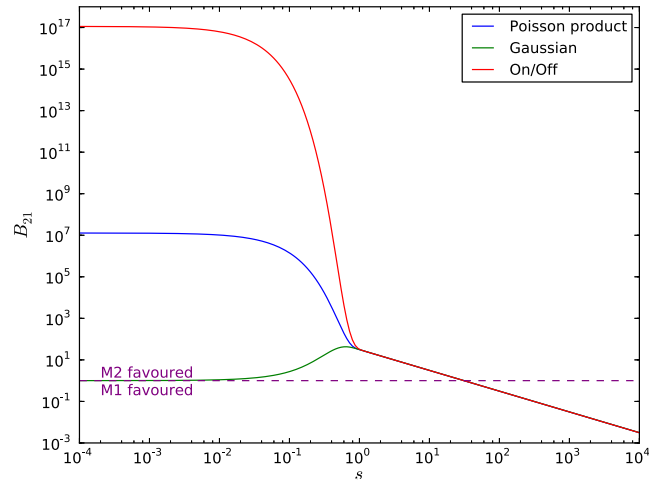


Figure B4. Dependence of the Bayes factor on hyperparameter s for the cases of Gaussian likelihood, the Poisson product likelihood, and the On/Off likelihood.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.