universe

MDPI

Article

Recovering Gamma-Ray Burst Redshift Completeness Maps via Spherical Generalized Additive Models

Zsolt Bagoly [1,2,*] and Istvan I. Racz [2]

[1] Department of Physics of Complex Systems, Eötvös Loránd University, H-1053 Budapest, Hungary
[2] Department of Natural Sciences, University of Public Services, H-1441 Budapest, Hungary; racz.istvan@uni-nke.hu
* Correspondence: zsolt.bagoly@elte.hu or bagoly.zsolt@uni-nke.hu

Abstract

We present an advanced statistical framework for estimating the relative intensity of astrophysical event distributions (e.g., Gamma-Ray Bursts, GRBs) on the sky tofacilitate population studies and large-scale structure analysis. In contrast to the traditional approach based on the ratio of Kernel Density Estimation (KDE), which is characterized by numerical instability and bandwidth sensitivity, this work applies a logistic regression embedded in a Bayesian framework to directly model selection effects. It reformulates the problem as a logistic regression task within a Generalized Additive Model (GAM) framework, utilizing isotropic Splines on the Sphere (SOS) to map the conditional probability of redshift measurement. The model complexity and smoothness are objectively optimized using Restricted Maximum Likelihood (REML) and the Akaike Information Criterion (AIC), ensuring a data-driven bias-variance trade-off. We benchmark this approach against an Adaptive Kernel Density Estimator (AKDE) using von Mises–Fisher kernels and Abramson's square root law. The comparative analysis reveals strong statistical evidence in favor of this Preconditioned (Precon) Estimator, yielding a log-likelihood improvement of $\Delta\mathcal{L} \approx 74.3$ (Bayes factor $> 10^{30}$) over the adaptive method. We show that this Precon Estimator acts as a spectral bandwidth extender, effectively decoupling the wideband exposure map from the narrowband selection efficiency. This provides a tool for cosmologists to recover high-frequency structural features—such as the sharp cutoffs—that are mathematically irresolvable by direct density estimators due to the bandwidth limitation inherent in sparse samples. The methodology ensures that reconstructions of the cosmic web are stable against Poisson noise and consistent with observational constraints.

Keywords: gamma-ray burst: general; methods: statistical; methods: data analysis; galaxies: distances and redshifts; cosmology: observations

Check for updates

Academic Editors: Yongfeng Huang and Patrizia Romano

Received: 12 December 2025
Revised: 10 January 2026
Accepted: 21 January 2026
Published: 24 January 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

In high-energy astrophysics and galactic astronomy, the analysis of the directional distribution of events (e.g., cosmic rays, neutrino sources, or specific stellar populations) requires rigorous non-parametric density estimation [1]. The analysis of cosmological data poses a fundamental epistemological challenge: while theoretical frameworks describe the evolution of continuous physical spaces (e.g., the cosmic matter density space, $\rho(\mathbf{x})$), observational campaigns yield discrete point sets (e.g., galaxy coordinates) or pixel intensity maps. Modern cosmological surveys, such as the Sloan Digital Sky Survey, the Dark Energy Survey, and the next-generation missions such as Euclid, provide celestial catalogs of

unprecedented volume and depth. These datasets, while rich in information, remain fundamentally discrete-stochastic collections of tracers on the celestial sphere.

1.1. The Cosmological Principle and the Large-Scale Distribution of Gamma-Ray Bursts

The observation of GRBs yields a distribution of events on the whole sky. While the angular coordinates are usually determined immediately upon detection, the event's cosmological redshift, z, is measured for only a subset, and after the event. These redshift measurements are obtained via two spectroscopic channels: observation of the optical afterglow or the spectroscopic analysis of the host galaxy. A fundamental requirement for any rigorous angular analysis of these sources is the determination of the sky exposure function, which gives the non-uniform detection probability across the sky. Here, we aim to reconstruct this exposure function and we also seek to determine whether separating the GRB population with measured redshift into distinct subsets based on the measurement method (afterglow-derived versus host-derived) yields a statistically superior model compared to a unified analysis. This allows us to assess if the selection effects inherent to these two methods favor a split treatment (Split Analysis) or if the samples are statistically compatible a for unified analysis (Lumped Analysis).

Due to their extraordinary luminosity, gamma-ray bursts serve as a powerful tool for studying distant regions and large-scale structures in the Universe [2–5]. Since they can be observed across vast cosmological distances, they provide a unique opportunity to study the distribution of matter and cosmic structures, as well as the physical conditions and evolution of galaxies in the early Universe. This includes analyzing the metal content and gas concentration of the interstellar medium (ISM) in their host galaxies [4,6]. Functioning as cosmological beacons, GRBs thus provide critical information about the evolution of the Universe in terms of star formation processes [7] and large-scale environmental effects. However, any cosmological inference drawn from GRB distributions—such as the evolution of the luminosity function or the cosmic event rate—is critically sensitive to observational selection effects. Recent studies have demonstrated that neglecting redshift incompleteness can lead to significant systematic biases in the reconstructed star formation rate and luminosity evolution [8–10]. Therefore, reconstructing the selection function (or completeness map) using a mathematically rigorous and numerically stable framework is not just a statistical task but a basic prerequisite for precision cosmology with GRBs.

The Cosmological Principle states that the Universe is homogeneous and isotropic on a large scale. This principle theoretically limits the largest size within which structures could have formed without violating the age of the Universe. This limit is generally set at around 260 h^{-1} Mpc [11]. However, precision observations in recent years have identified an increasing number of structures arising from the clustering of bright astronomical objects that exceed this theoretical limit. These challenging structures include large-scale features traced by gamma-ray bursts, such as the Giant GRB Ring [12,13] and the Hercules-Corona Borealis Great Wall (HerCrbGW) [14,15]. As the largest known structure, with an estimated diameter of 2.5–3 Gpc, the HerCrbGW raises serious questions about its compatibility with the Cosmological Principle, underscoring the need for a closer connection between large-scale cosmological models and local observations. The HerCrbGW was initially discovered in a sample of 283 known redshifted GRBs in the range $1.6 \leq z \leq 2.1$ and was later confirmed and extended [16,17]. However, since GRBs are transient events and their detection rate is strongly influenced by selection effects, the true size and shape of these large-scale structures remain uncertain [18]. Overcoming these uncertainties is critical and requires robust, non-parametric methodologies for density estimation, such as those presented in this article. Previous attempts to characterize these distributions have relied on density-ratio methods or geometric tessellations, which often introduce artifacts especially

in the low-exposure regions. Our goal is to create a framework that preserves structural fidelity without the strict bandwidth limitations imposed by sparse datasets.

1.2. The Problem of Sky Coverage Completeness for Redshifted GRBs

The observed angular distribution of GRBs is not uniform, but is modulated by a complex, varying 'sky exposure function' [19]. This function represents a combined probability density that depends on the celestial mechanics and technical specifications of the detecting satellite (e.g., detector sensitivity, field of view, and pointing history). Accurately reconstructing this modulator is essential for distinguishing between instrumental non-uniformity and clustering of matter on cosmological scales. Furthermore, the probability of detecting optical afterglow, a prerequisite for determining redshift and detailed characterization, is governed by logistical factors related to the ground-based tracking network. These factors include telescope availability, local meteorological conditions, lunar phase, and human-induced variables such as academic schedules or maintenance periods. Consequently, the effective selection function is a complex surface probability density derived from the convolution of satellite dynamics and ground constraints.

Let $\mathcal{D}_{total}$ be the complete catalog of detected bursts (e.g., from Swift BAT), and $\mathcal{D}_z \subset \mathcal{D}_{total}$ be the subset with spectroscopically confirmed redshifts [20]. Our goal is to identify the regions exhibiting a statistically significant relative excess or deficiency of redshift measurements on the sky.

Adaptive geometric methods like Voronoi Diagram Field Estimator (VDFE) [21] and Delaunay Tessellation Field Estimator (DTFE) [22] are standard for cosmic web analysis in surveys such as DESI [23] and Euclid [24]; the probabilistic reconstruction of the relative exposure function presented here offers a superior framework for handling complex selection functions in the GRB population.

While the VDFE and DTFE are standards, they were excluded from the primary comparison in favor of Kernel Density Estimation (KDE) for GRB sky exposure functions [25]. Simple but optimized Gaussian density estimators demonstrated that both VDFE and DTFE are prone to generating spurious local maxima, or 'bogus hot spots', particularly in regions like the North Galactic Pole. To recover a physically realistic detection probability function, these geometric estimators require an additional stage of kernel smoothing to suppress these artifacts. In contrast, it was shown that KDE yields a self-consistent reconstruction of the density field without introducing significant artificial features or requiring secondary post-processing. Therefore, the KDE method was selected as the more robust density-based baseline for this study too.

Cosmologists and researchers working on large-scale studies are the target audience for this work. By implementing a Bayesian-driven preconditioning technique this method recovers some high-frequency structural features in the selection function that were previously lost, exceeding the bandwidth constraints of conventional estimators. It can be a useful tool for the catalogs of the new high-redshift missions (e.g., SVOM [26], Einstein Probe [27] and Theseus [28]), where it is critical to differentiate between physical clustering and observational bias.

2. Data Description

Here, we use a comprehensive database of the Swift GRBs with measured positions and spectroscopic redshifts. The Swift GRB Table[1] is a public catalog of all Swift's GRB observations. It contains the most γ parameters, such as the positional information, the different durations, fluxes and fluences of a GRB event [29]. Several X-ray and UV-optical parameters and comments from all three Swift instruments are also included. Additionally, the Swift-XRT GRB Catalogue[2] from the UK Swift Science Data Centre (UKSSDC) contains

all published XRT light curves and spectra with XRT positions. The data was accessed on 17 November 2025, with GRB 251112A beeing the last event included.

The Swift mission employs a hierarchical positioning strategy initiated by the Burst Alert Telescope (BAT), a wide-field coded-aperture instrument operating in the 15–150 keV range [30]. Upon triggering, the BAT reconstructs sky images to provide an initial localization with an accuracy of 1–4 arcminutes (90% confidence) within seconds [31]. Once the spacecraft slews to the BAT target, the XRT position repository, maintained by the UK Swift Science Data Centre, generates the most accurate astrometric localizations. The primary data product, known as the 'Enhanced Position', utilizes a sophisticated cross-correlation technique to minimize systematic errors inherent in the spacecraft's pointing [32]. It significantly reduces the radial systematic error to approximately 1.4 arcseconds (90% confidence) [33]. Here, we use this enhanced positional information, if available for a given event; otherwise, we use the Swift BAT position. On the figures, like Figure 1, black points show the actual sky positions of the GRB dataset.

The redshift dataset was constructed using GCN data and was cross-checked using Jochen Greiner's GRB Big Table[3], publications and the Gamma-Ray Burst Online Index (GRBOX)[4]. We used here the same spectroscopic redshift dataset used earlier by [16,34,35], with redshift observations corrected and updated up until 31 August 2022 [17]. The list of the events was limited to agree with the Swift observations.

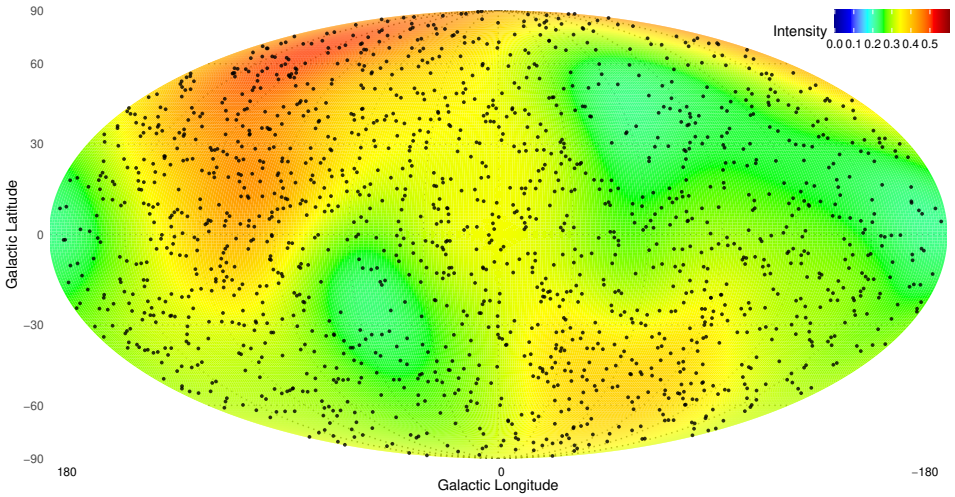


Figure 1. Adaptive KDE estimation of the entire population according to Abramson [36]. This map serves as the basis for the structural constraints of the Preconditioned model. The black points show the actual GRBs' sky positions. They form the dataset for the AKDE producing the density map. The intensities are non-normalized relative values.

We explicitly excluded photometric redshifts and estimates based on limits (e.g., Lyα limits), as these data tend to introduce significant uncertainties in radial distance. Since our final goal is the investigation of the full spatial distribution, we restricted our analysis to precise spectroscopic redshifts only.

The selection results in a total of 1665 GRBs with Swift positional information (BAT positions and, where available, XRT enhanced positions). Of this sample, 427 GRBs possess precise spectroscopic redshifts, with 300 derived from afterglow observations and 127 from host galaxy measurements.

3. The Kernel Density Estimation

The Kernel Density Estimation is a non-parametric alternative for density reconstruction [37]. In the classical approach, two independent intensity functions, $\lambda_{total}(\mathbf{x})$ and

$\lambda_z(\mathbf{x})$, are estimated using KDE. The angular relative probability (or redshift completeness), denoted as $R(\mathbf{x})$, is then derived by calculating their ratio:

$$R(\mathbf{x}) = \frac{\hat{\lambda}_z(\mathbf{x})}{\hat{\lambda}_{total}(\mathbf{x})} \tag{1}$$

In this naive approach, we determine relative probability by dividing the density estimate of the redshifted subset ($\mathcal{D}_z$) by the density estimate of the total observed population ($\mathcal{D}_{\text{total}}$), applying standard KDE to both datasets independently.

Applying the KDE method poses significant statistical and numerical challenges [38,39]. Unlike parametric models, which assume a specific underlying shape (e.g., a single cluster), non-parametric methods allow the data to determine the complexity of the angular structure, but standard KDE on the sphere often suffers from the classic bias-variance trade-off. A fixed bandwidth tends to oversmooth dense clusters (erasing high-frequency angular information) or undersmooth sparse background regions (introducing spurious noise). Insufficient bandwidth ($h \to 0$) results in an undersmoothed, high-variance estimate that is sensitive to the stochastic placement of individual points (overfitting). In contrast, excessive bandwidth ($h \to \infty$) produces an oversmoothed, high-bias estimate in which significant structural details are obscured. This is a classic example of the bias-variance trade-off known in statistical learning [40]. Although standard KDE approaches can be used, they introduce the aforementioned bandwidth dilemma [41]: an inadequately optimized bandwidth can result in the deletion of true large-scale anisotropies (oversmoothing) or the generation of false, noise-driven peaks (undersmoothing).

The most critical problem with this ratio-based approach is the singular behavior: in areas where the background density of $\lambda_{total}(\mathbf{x})$ approaches zero (e.g., zones obscured by the Galactic plane), the ratio tends to diverge asymptotically to infinity, causing extreme variance and meaningless artifacts (spikes) in the resulting map. Furthermore, the error propagation is significant, as the biases and variances in two independent estimators accumulate during the division process, often amplifying Poisson noise. Finally, the bandwidth selection dilemma complicates the analysis, as the two density functions may require different optimal smoothing parameters, the coordination of which poses a non-trivial optimization problem for ratio formation.

To solve this problem, we apply the adaptive method of Abramson [36], adapted to spherical geometry [42], and introduce a robust Bayesian logistic regression framework (GAM/SOS), which is detailed below.

Adaptive Spherical KDE Pilot Estimation and Bandwidth Selection

One way to solve density estimation problems is to use adaptive bandwidth. This method varies the kernel width in inverse proportion to the square root of the local density, which reduces the asymptotic bias of the estimation function from $\mathcal{O}(h^2)$ to $\mathcal{O}(h^4)$.

For a unit vector $\mathbf{x}$ and a mean vector $\boldsymbol{\mu}$, the spherical counterpart of the Gaussian distribution is the Von Mises–Fisher distribution [43], whose probability density function is given by:

$$f_{vMF}(\mathbf{x}; \boldsymbol{\mu}, \kappa) = C(\kappa) \exp(\kappa \boldsymbol{\mu}^T \mathbf{x}) \tag{2}$$

where $\kappa \geq 0$ is the concentration parameter (which is analogous to the inverse variance, $\kappa \sim 1/\sigma^2$). The normalization constant $C(\kappa)$ is critical for strict likelihood estimation [44] and is defined as follows:

$$C(\kappa) = \frac{\kappa}{4\pi \sinh(\kappa)} \tag{3}$$

In the first step, a pilot density estimate, $\tilde{f}(\mathbf{x})$, is created using a fixed global concentration parameter κ_{pilot}. To determine the κ_{pilot}, we use Likelihood Cross-Validation,

also known as Maximum Likelihood Cross-Validation. This procedure, which was established by Stone [45] in statistical prediction and standardized by Silverman [37] for density estimation, ensures data-driven bandwidth selection. The specific extension of the method to spherical data and directional distributions, as well as the optimization of the concentration parameter (κ), is discussed in detail in the work of Hall et al. [42]. We define the Leave-One-Out estimator for the density of data point $\mathbf{X}_i$ using all other points $j \neq i$:

$$\hat{f}_{-i}(\mathbf{X}_i; \kappa) = \frac{1}{N-1} \sum_{j \neq i} C(\kappa) \exp(\kappa \mathbf{X}_j^T \mathbf{X}_i) \tag{4}$$

The optimal pilot parameter is obtained by maximizing the log-likelihood of the data:

$$\kappa_{pilot} = \arg\max_{\kappa} \sum_{i=1}^{N} \log \hat{f}_{-i}(\mathbf{X}_i; \kappa) \tag{5}$$

This optimization is performed numerically, without constraints, allowing the data to objectively dictate the underlying smoothness.

Using the optimized $\tilde{f}(\mathbf{x})$ pilot estimate, we calculate the local bandwidth modifiers (λ_i) for each data point according to Abramson's square root law [36]:

$$\lambda_i = \left(\frac{\tilde{f}(\mathbf{X}_i)}{g} \right)^{-1/2} \tag{6}$$

where g is the geometric mean of the pilot densities, which serves as a scaling factor. In areas where the pilot density $\tilde{f}(\mathbf{X}_i)$ is high (clustering), $\lambda_i < 1$, which prescribes a narrower kernel function in order to preserve detail. In areas of low density, $\lambda_i > 1$, which prescribes a wider kernel function to smooth the variance. Since the concentration of vMF κ is inversely proportional to the square of the bandwidth ($h^2 \propto 1/\kappa$), the derivation of the local concentration parameter κ_i associated with the i-th event is as follows:

$$\kappa_i = \frac{\kappa_{pilot}}{\lambda_i^2} \tag{7}$$

The final adaptive $\hat{f}(\mathbf{x})$ probability density estimate is the normalized sum of the variable kernel functions at any arbitrary point $\mathbf{x}$ on the sphere:

$$\hat{f}(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^{N} C(\kappa_i) \exp(\kappa_i \mathbf{X}_i^T \mathbf{x}) \tag{8}$$

Here, $C(\kappa_i)$ ensures that the integral of each kernel function on the sphere is unity, guaranteeing that the total estimate $\hat{f}(\mathbf{x})$ is a valid PDF:

$$\int_{S^2} \hat{f}(\mathbf{x}) \, d\Omega = 1 \tag{9}$$

4. Methodology: Bayesian Preconditioning via Spherical GAMs

To address the inherent instabilities of the traditional density ratio approach, we reformulate the completeness problem as a conditional probability estimation task. Instead of independently estimating the generative densities for the total and redshift samples—which leads to divergence and bandwidth mismatch—we utilize a statistical framework. See Appendix B for a detailed mathematical derivation of the Restricted Maximum Likelihood (REML) optimization and the Splines on the Sphere (SOS) basis functions.

4.1. The Bayesian Approach: High-Statistics Priors

Using the entire Swift BAT GRB sample ($N_{total} = 1665$) as an informative spatial prior is our main innovation. For every detection, we establish a binary indicator $y_i \in \{0, 1\}$ that indicates if a measured redshift is present. We aim to recover the latent probability surface $p(\mathbf{x}) = P(z > 0|\mathbf{x})$. By embedding this in a Generalized Additive Model (GAM) with a logit link function, we can guarantee that the resulting efficiency map is physically consistent and rigorously constrained within $[0, 1]$. This method prevents the phantom leaks and spikes characteristic of Adaptive KDE ratios by forcing the model to satisfy the geometric restrictions of the survey (e.g., low probability areas) inherent in the entire sample.

4.2. Structural Optimization

To provide an isotropic, pole-singularity-free depiction of the celestial sphere, we use Splines on the Sphere. SOS enables continuous interpolation and sub-pixel resolution, in contrast to grid-based techniques. The Akaike Information Criterion (AIC) and Restricted Maximum Likelihood objectively control the model's smoothness and complexity, providing a data-driven bias-variance trade-off.

For the full Lumped dataset, the AIC selects an optimal base dimension of $k = 44$ (Figure 2). This parameter sets the number of basis function nodes in the SOS framework, allowing the model to accurately capture the distribution without overfitting.

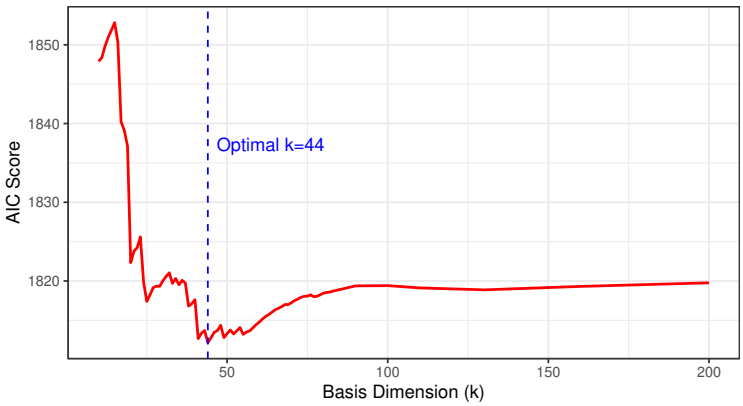


Figure 2. Model selection based on the Akaike Information Criterion on the lumped dataset. The optimal base dimension is found at $k = 44$, where the AIC value reaches a global minimum. This parameter determines the upper limit of the smoothness of the spherical splines.

5. Statistical Analysis

To determine whether the proposed Bayesian methodology provides a more accurate representation of physical reality, we perform a comparative analysis between the two approaches with respect to the signal population (events with redshift). One of the models examined is M_{KDE}, a traditional Adaptive Kernel Density Estimator (AKDE), which is derived exclusively from the distribution of signal events (N_z) (Figure 1). This contrasts with the M_{Precon} model, a complex Preconditioned Estimator, defined as the product of the background density ρ_{total} and the conditional probability $P(\text{Redshift}|x)$ derived using Bayesian Generalized Additive Models (Figure 3):

$$\hat{f}_{Precon}(x) \propto P(\text{Redshift}|x) \cdot \rho_{total}(x) \tag{10}$$

We use two different statistical frameworks for the objective evaluation of the realism and performance of the models: Likelihood and Discrimination Analysis (ROC/AUC).

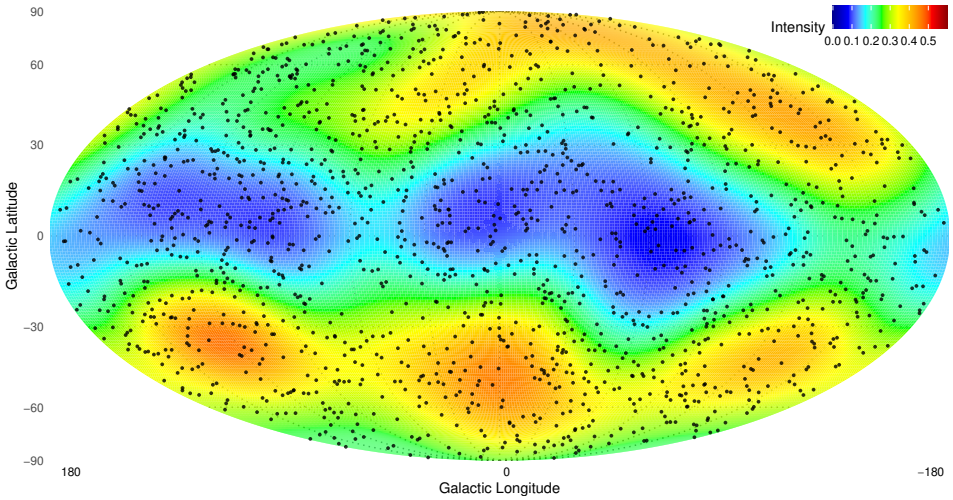


Figure 3. Bayesian relative probability map, $P(\text{Redshift}|\mathbf{x})$ is the conditional probability distribution estimated by the logistic GAM model. The map shows the spatial probability of the redshift measurement (signal) relative to the GRB detection (background). The black dots marks the GRB events' angular positions. It uses the optimal basis dimension $k = 44$ selected by the AIC value (see Figure 2).

5.1. Likelihood Analysis

The likelihood principle provides a rigorous and objective measure for comparing probability density functions [37,45]. To scientifically prove that the Precon distribution is superior to the standard AKDE, we must demonstrate that it minimizes information loss, i.e., it maximizes the predictive likelihood of the true distribution of the signal.

We compare the log-likelihood ($\mathcal{L}$) of the observed GRB events for both density maps. We define the log-likelihood of the $\hat{f}$ model for the signal data $X_z = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ as follows:

$$\mathcal{L}(\hat{f}|X_z) = \sum_{i=1}^{n} \ln\left(\hat{f}(\mathbf{x}_i)\right) \tag{11}$$

A higher $\mathcal{L}$ value indicates that the model has correctly concentrated a significant portion of the probability mass on the areas where the events actually occur. This is a measure of calibration and precision. To compare the two models, we calculate the likelihood ratio (or Bayes factor proxy):

$$\Delta\mathcal{L} = \mathcal{L}_{Precon} - \mathcal{L}_{KDE} \tag{12}$$

We interpret this difference based on the scale of Kass and Raftery [46] ($2\Delta\mathcal{L}$), where the range $0 \ldots 2$ is negligible, $2 \ldots 6$ is positive, $6 \ldots 10$ is strong, and $2\ Delta\mathcal{L} > 10$ represents very strong (decisive) evidence in favor of the model.

5.2. Discrimination Analysis (ROC/AUC)

Since the goal is to distinguish the signal (GRBs with redshift) from the rest of the distribution, in the absence of true non-events, we applied the Presence-Background ROC analysis methodology [47]. To evaluate the discriminative power of the density estimators, we generated $N = 12{,}288$ HEALPix grid points uniformly distributed on the sphere, which served as pseudo-negative class (background).

The Area Under the Curve, which quantifies performance, represents the probability that a randomly selected signal event has a higher estimated density than a randomly selected background point:

$$AUC = P(\hat{f}(x_{signal}) > \hat{f}(x_{noise})) \tag{13}$$

It is important to note that the AUC is a rank-order metric; so, it is invariant with respect to the absolute values of density [48]. If a model physically underestimates probabilities but preserves the order, it can still achieve a high AUC value [1].

6. Results and Discussion

6.1. Results of Statistical Comparative Analysis

We evaluated the Information Criteria (Log-Likelihood) and Discriminative Power (ROC/AUC) of the proposed Precon model against the standard AKDE. During the empirical evaluation of the datasets, the goodness-of-fit analysis (Likelihood) revealed overwhelming evidence in favor of the Precon Estimator (Figure 4). The $\Delta\mathcal{L} \approx 74.3$ Log-Likelihood difference ($\mathcal{L}_{Precon} \approx -519.3$ vs. $\mathcal{L}_{KDE} \approx -593.6$), which corresponds to a standard evidence score of $2\Delta\mathcal{L} \approx 148.6$, represents a Bayes factor of $K > 10^{30}$. This indicates that the Precon distribution provides a significantly more accurate description of the event generation process, minimizing information loss.

In contrast, in terms of discrimination analysis (ROC), i.e., signal-to-noise separation, both models performed similarly (Figure 5). AKDE achieved a marginally higher Area Under the Curve ($AUC_{KDE} \approx 0.682$) compared to the Precon model ($AUC_{Precon} \approx 0.674$). This difference of less than 1.2% is negligible and is likely due to the sensitivity of ranking metrics to local noise spikes.

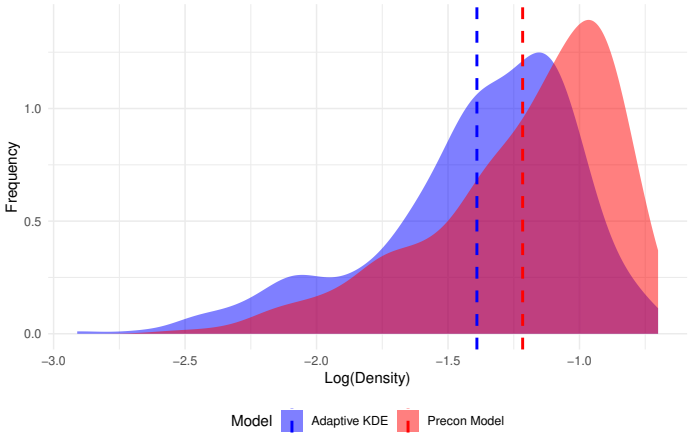


Figure 4. Goodness of fit. Distribution of log-likelihood values at actual event positions. The distribution of the Precon model shifted significantly to the right compared to the KDE model. The log-likelihood difference of $\Delta\mathcal{L} \approx 74.3$ provides decisive evidence for the superiority of the Precon model.

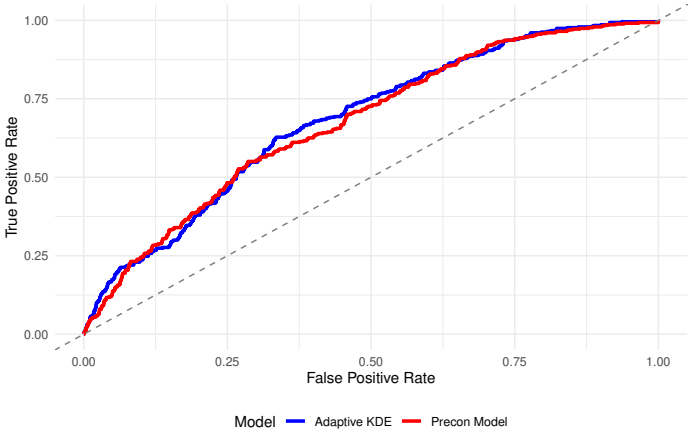


Figure 5. Discrimination analysis (ROC) on the lumped dataset for the AKDE (blue) and Precon (red) models. Based on a comparison of AUC values, the discriminatory power of the two models is statistically similar according to ranking metrics.

The Duality of Matching and Ranking

Based on the above, we can observe an apparent contradiction: the Precon model is significantly better in terms of likelihood ($\Delta\mathcal{L} \gg 10$), but marginally weaker or equivalent in terms of discrimination ($\Delta AUC \approx 0$). This phenomenon can be explained by the fundamental difference between fitting and ranking.

AKDE, because it comes from a small sample (N_z), tends to overfit local variance, often creating narrow 'spikes' (high variance) around observed redshift events [49]. On the one hand, in the case of AUC, KDE places the spikes exactly on the training points, thus ranking these points higher than the background, which artificially inflates the ranking-based metrics. On the other hand, in the case of Likelihood, although KDE fits well to specific points, it often fails to model the distribution shape between points. It may assign a probability close to zero to valid regions in the immediate vicinity of events (undersmoothing). In Likelihood calculations, an estimated probability of nearly zero for a real event results in a huge mathematical penalty ($\ln(0) \to -\infty$), which causes a significant disadvantage in the fit test compared to the Precon model.

By conditioning the signal probability $P(\text{Redshift}|x)$ on the high-statistic total density ρ_{total}, the Precon method incorporates structural information from the background that is missing when analyzing isolated subsets with small sample sizes [1]. This effectively acts as a Bayesian prior, stabilizing the density estimation and better handling the bias-variance trade-off.

6.2. Subgroup Analysis and Validation of the Split Model

As a supplement to the lumped model analysis, we separately examined two specific subgroups of the signal population (afterglow and host galaxy based redshifts) and evaluated the effectiveness of this model separation compared to the AKDE approach.

In the dataset with afterglow based redshifts (optimal $k = 41$), the Precon model retained its statistical superiority (top of Figure 6). The likelihood analysis resulted in a value of $\Delta\mathcal{L} \approx 16.11$ in favor of Precon (Precon: -361.54, KDE: -377.65), which is very strong evidence on the Kass–Raftery scale (Section 5.1). In terms of discriminator power, the two models performed similarly ($AUC_{Precon} \approx 0.65$, $AUC_{KDE} \approx 0.67$).

For host-galaxy-based redshifts ($k = 23$, Figure 7), which comprise a smaller sample size, the likelihood difference decreases significantly ($\Delta\mathcal{L} \approx 0.29$, Figure 6, center). Consequently, this model imposes a more stringent penalty on complexity to avoid misinterpreting stochastic Poisson noise as genuine physical clustering.

Here, the advantage of the Precon model is negligible, while the AUC values showed the superiority of KDE ($AUC_{Precon} \approx 0.63$, $AUC_{KDE} \approx 0.70$). This suggests that for data with lower information content, the structural constraints of Precon are less effective.

Finally, to verify model complexity, we performed a Split vs. Lumped hypothesis test. We compared the lumped model with the split model, where the density was the weighted sum of the two sub-models (separate datasets for the afterglow and host galaxy based redshifts). The result (bottom of Figure 6) of the likelihood analysis shows that the lumped model is statistically superior ($\mathcal{L}_{Lumped} \approx -1599.44$, $\mathcal{L}_{Split} \approx -1615.57$, $\Delta\mathcal{L}_{diff} \approx -16.13$). The afterglow-host galaxy partitioning added redundant complexity, resulting in overfitting without improving the model's accuracy in describing the real distribution.

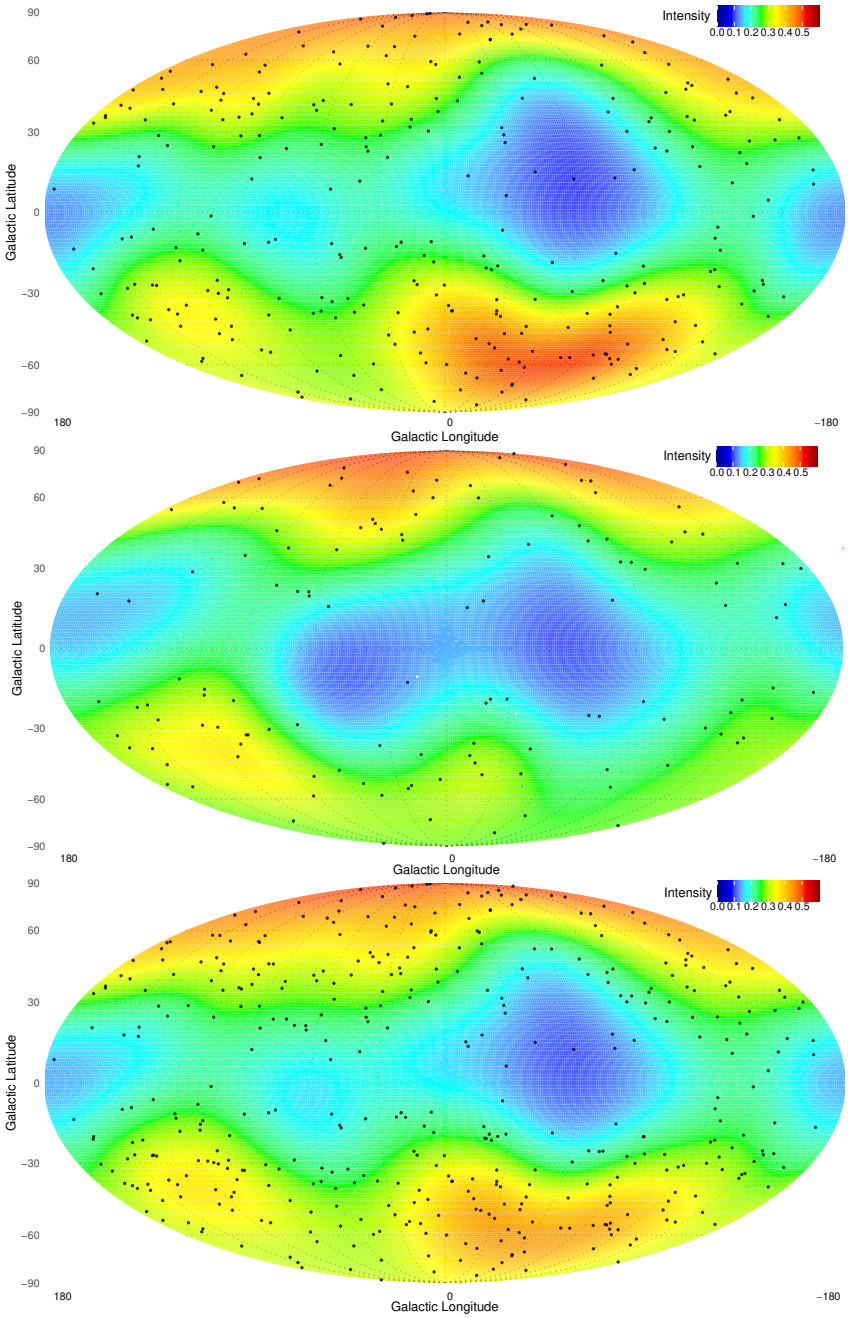


Figure 6. Precon density estimate for the different groups. The black dots marks the GRB events' angular positions. The intensities are non-normalized relative values, but the maximums are the same. (Top) afterglow determined redshift subgroup: based on the likelihood analysis, this model statistically significantly ($\Delta\mathcal{L} \approx 16.1$) outperforms the KDE. (Center) host-galaxy-determined redshift subgroup: The likelihood ratio ($\Delta\mathcal{L} \approx 0.29$) here indicates only a marginal advantage for Precon, reflecting the weaker relationship between the model and the data in this subgroup. (Bottom) superposition of the two density estimates of the two independent submodels (afterglow and host galaxy) forms the Split Density Map. The statistical comparison shows that the Lumped model outperforms the Split model ($\Delta\mathcal{L}_{Split} \approx -16.13$); thus, it is suboptimal (Figure 8). A comparison of these figures reveals that the afterglow-derived map shows a high-density region in the lower left region, a feature absent in the host-derived map. Furthermore, the under-density void at $(l \approx 60^\circ, b \approx 0^\circ)$ in the host-galaxy map is less pronounced in the (combined) Split Density map. Due to the disparity in sample sizes ($N_{\text{afterglow}} = 300$ versus $N_{\text{host}} = 127$), the result map is primarily governed by the afterglow distribution.

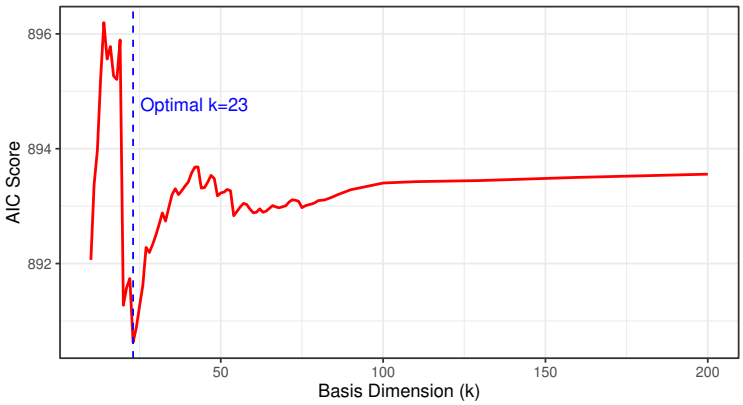


Figure 7. Model selection for the host galaxy determined redshift subgroup. The minimum of the AIC curve is found at a lower dimension, $k = 23$. This more economical model suggests that this dataset contains less data points and less detailed spatial information.

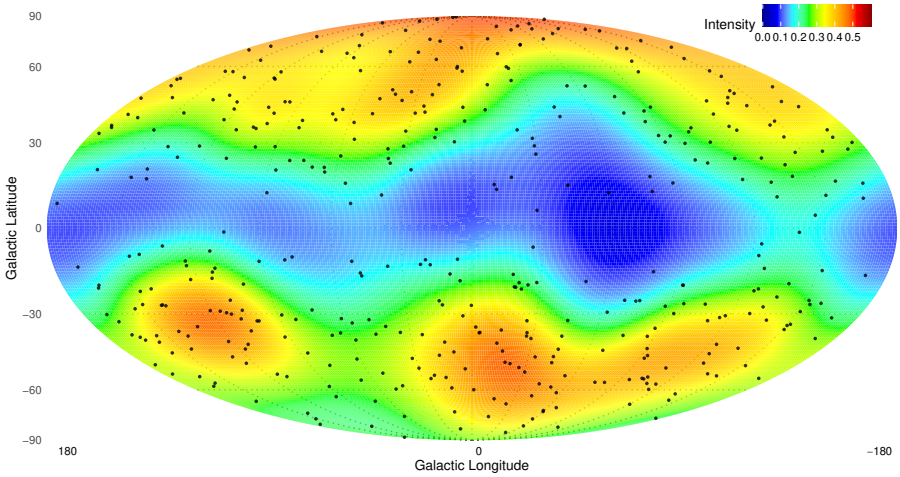


Figure 8. Density estimate using Precon map and Lumped model. This map is proved to be statistically better to the Split model.

6.3. Pixel-to-Pixel Correlation Analysis

To check the concordance between the two estimators and to identify their behaviors across distinct signal regimes, a pixel-by-pixel correlation analysis of the probability density functions was performed. Both the Precon and AKDE intensity fields were normalized such that their total probability mass equaled unity, thereby ensuring a comparison independent of arbitrary scaling factors.

The resulting correlation density plot (Figure 9) reveals three distinct regimes of model behavior driven by the conservation of probability mass.

In the Low-Density Regime ($\lesssim 2.0 \times 10^{-5}$), a significant asymmetry is visible where the distribution deviates above the identity line. This indicates that the AKDE estimator consistently assigns higher probability to the low-probability (void) regions than the Precon model. This is a characteristic signature of boundary bias: whereas the Precon model successfully fits the sharp cutoff of the Galactic Disc, the AKDE is spatially bounded by its kernel width, causing probability mass to 'spill' into unobserved regions.

In the Intermediate Regime ($2.0 - 4.0 \times 10^{-5}$), to compensate for the excess mass assigned to the voids and peaks by the AKDE, the Precon model concentrates slightly more mass in the intermediate signal domain. The grid points cluster tightly along the diagonal (Pearson's $r \approx 0.974$), demonstrating that Precon preserves the general signal structure of the standard estimators.

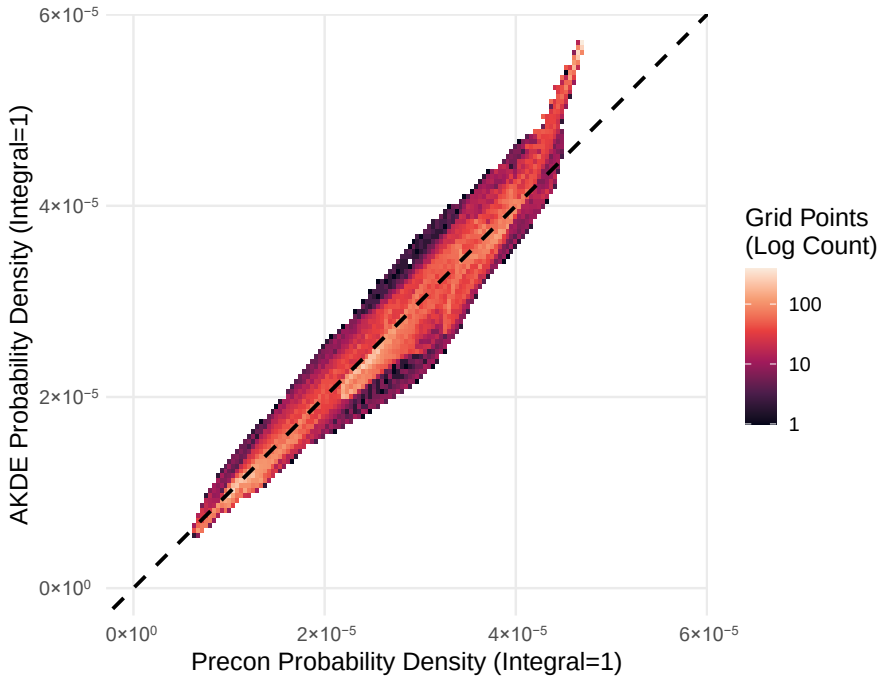


Figure 9. The logarithmic density of grid points within the Precon-AKDE probability plane is displayed, wherein both intensity fields have been normalized to a unit spherical sum. A significant skew of the distribution above the black identity line is observable at low values. This asymmetry demonstrates that density within voids is systematically overestimated by the AKDE relative to the Precon model due to kernel smoothing limits.

In the latex High-Intensity Regime (>4.0×10^{-5}), a slight scatter and upward deviation is observable. This reflects the different regularization strategies: the Abramson adaptive bandwidth ($h \propto \hat{f}^{-1/2}$) allows AKDE to fit narrow, high-amplitude spikes (potentially overfitting Poisson noise), whereas the Precon model is governed by a global smoothness penalty (REML), resulting in slightly more conservative peak estimates.

6.4. Spatial Artifact Analysis: Spurious Spikes and Signal Leakage

We conducted a point-by-point difference analysis of the probability density fields in order to quantify the structural limitations of the Adaptive Kernel Density Estimator in comparison to the Precon model. Three classes can be seen in the resulting difference map (Figure 10). The map is punctuated by sharp, high-amplitude areas of excess density, where $\hat{f}_{AKDE} \gg \hat{f}_{Precon}$. These correspond to the spurious spikes where the AKDE adaptive bandwidth, scaling as $h \propto \hat{f}^{-1/2}$, narrows excessively in high-density regions. The Galactic central region appears to be covered by a wide, diffuse region of excess probability. The Signal Leakage artifact is demonstrated by this. The AKDE kernel width grows unduly in data-sparse zones to make up for the absence of local support. As a result, probability mass from high-latitude sources is artificially dispersed throughout the Galactic Disk. Surrounding these features, large void regions appear in Cyan/Blue, indicating where the AKDE density is lower than the Precon estimate ($\hat{f}_{AKDE} < \hat{f}_{Precon}$). This shows the overall cost of the unphysical spikes. The estimator must decrease the background density elsewhere in order to meet the normalization constraint because these spikes take up such a significant portion of the overall probability.

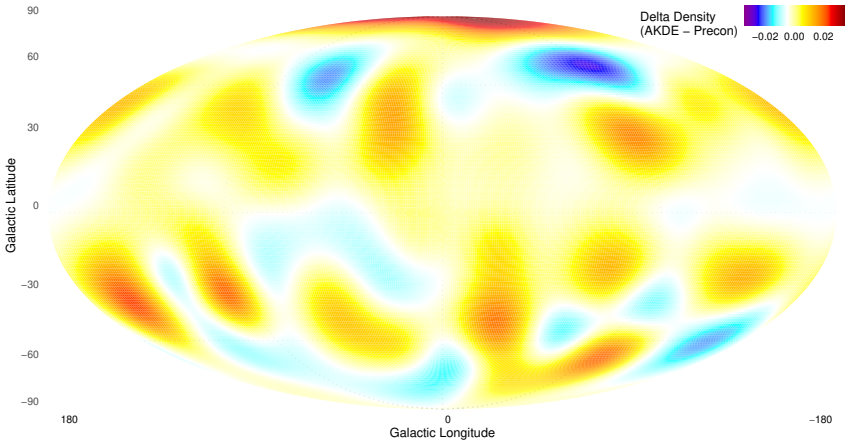


Figure 10. Spatial artifact analysis of the GRB redshift completeness maps: the $\Delta = \hat{f}_{AKDE} - \hat{f}_{Precon}$ difference between the normalized AKDE and Preconditioned probability densities is displayed. The yellow/red regions ($\Delta > 0$) identify AKDE artifacts, which are discrete, high-amplitude spurious spikes caused by overfitting (red islands) and diffuse signal leakage where the kernel smears mass into the Galactic Disk (broad yellow zones). The cyan/blue regions ($\Delta < 0$) show background depletion areas, where the AKDE systematically underestimates the density to compensate for the probability mass consumed by the spikes.

6.5. Limitations and Physical Biases

It is important to note that the current analysis is based on the Swift BAT catalog, which is inherently biased towards harder X-ray energies and dominated by long gamma-ray bursts. Short GRBs, typically originating from neutron star mergers, are probably under-represented in redshift catalogs due to their fainter optical afterglows and different host galaxy environments. Furthermore, missions operating in different energy bands, such as the new SVOM or Einstein Probe, will probe sligthly different populations of the GRB.

However, the distinction between these GRB populations is not only a matter of intrinsic physics but is also significantly influenced by instrumental effects. Comparative studies between Swift BAT and Fermi GBM have shown that for the exact same set of jointly detected GRBs, the measured durations (T_{90}) systematically differ [50]: Fermi GBM tends to measure longer durations for short GRBs, while Swift BAT measures longer durations for long ones. This instrumental dependency is further highlighted by the fact that the T_{90} distribution of the same GRB sample requires three lognormal components when observed by Swift, but only two when observed by Fermi. These discrepancies suggest that the traditional classification into short and long events is instrument-specific and does not uniquely represent the underlying physical mechanisms. Consequently, while certain types like neutron star mergers might be under-represented in our current redshift-complete sample, the lack of a clear selection effect in the redshift measurement process itself—independent of these fuzzy classification boundaries—suggests that our completeness maps is a useful tool for characterizing the observational biases of the respective satellite missions.

The Bayesian GAM framework, however, is intended to be mission-neutral. Once sufficient statistics are available, the factorized estimation approach ($\hat{\lambda}_z = \hat{\lambda}_{total} \cdot \epsilon$) can be directly applied to any future catalog or particular sub-population (e.g., short GRBs only). Our approach objectively recovers the selection function for any given observational setup, providing the mathematical framework to account for these intrinsic biases. Recovering high-frequency features like sharp low probability areas is critical to prevent observational artifacts from being misidentified as genuine cosmological anisotropies. Accurate recovery minimizes signal leakage into zero-visibility regions, enabling the use of physically realistic exposure maps instead of those degraded by sparse sample constraints. These high-resolution maps are essential for rigorous large-scale isotropy investigations.

7. Summary and Conclusions

We concluded that the Precon Estimator constitutes a statistically improved framework for recovering the underlying angular distribution of GRBs with determined redshift compared to traditional adaptive kernel methods. The empirical analysis demonstrates that the Precon model yields a log-likelihood improvement of $\Delta\mathcal{L} \approx 74.3$ over the standard AKDE, which corresponds to a standard evidence score of $2\Delta\mathcal{L} \approx 148.6$, implying a Bayes factor exceeding 10^{30}. This illustrates that the Precon model offers a significantly more accurate probabilistic description of the event selection process. While the discrimination analysis suggests comparable performance ($\Delta\text{AUC} < 1.2\%$), this parity is attributable to the rank-order invariance in the AUC metric, which rewards the tendency of AKDE to overfit local noise spikes in sparse datasets without penalizing poor calibration. Conversely, the Precon Estimator minimizes information loss, providing a robust reconstruction of the physical intensity field that is stable against Poisson noise.

The theoretical efficacy of the Precon Estimator is derived from its ability to circumvent the fundamental bandwidth limitation that limits direct adaptive estimators. The Asymptotic Mean Squared Error analysis shows that the angular resolution of any direct estimator is asymptotically bound by the sample size of the signal population (n), scaling as $h \propto n^{-1/6}$. For the sparse redshifted GRB sample ($n \ll N$), this constraint forces the smoothing kernel to function as a low-pass filter with a soft ceiling on resolution, rendering it mathematically impossible to resolve high-frequency angular features. By decomposing the problem into a wideband exposure map and a narrowband selection efficiency, the method recovers structural information beyond the Nyquist–Shannon sampling limit of the sparse redshift subset. This ensures that the resulting cosmological maps remain physically consistent with global observational constraints.

The resulting high-fidelity redshift completeness maps have direct implications for both GRB population synthesis and observational cosmology. In population studies, the ability to assign a coordinate-dependent completeness weight to individual events allows for a more precise derivation of the intrinsic rate density and luminosity function. This approach effectively corrects for spatial spatial variance induced by observational constraints and satellite mechanics [19]. Furthermore, the model serves as a noise-filtered background for Large-Scale Structure studies and tests of the Cosmological Principle.

This methodology is directly applicable to data from current and upcoming missions, including SVOM, THESEUS, and the Einstein Probe. Implementing this Bayesian framework on early mission catalogs will facilitate the rapid identification of sky-dependent efficiencies, ensuring a more informed interpretation of the high-redshift Universe. While the empirical validation presented here uses the Swift-BAT catalog, the Precon framework is fundamentally mission-agnostic. Due to its generalizable nature, the technique can be applied to any observational setup in which the recovery of a sparse subset can be preconditioned using a high-statistics background.

From an information-theoretic perspective, the Precon approach represents a minimization of the generalized Kullback–Leibler divergence between the estimated and true intensity fields. By conditioning the signal probability on the high-statistics background density, the method injects essential structural constraints—effectively acting as informative priors—that prevent the 'probability leakage' into unobserved regions (such as the Zone of Avoidance) that characterizes standard KDE. While the direct AKDE maximizes entropy solely under the weak constraints of the sparse redshift sample, the Precon model incorporates side-channel information provided by the non-redshifted population. This yields a Maximum Entropy solution that is consistent with global observational constraints, validating the hypothesis that a factorized, physics-aware statistical framework is essential for the mapping of cosmological objects subject to complex selection functions.

Author Contributions: Conceptualization, all; data curation, all; formal analysis, all; Investigation, all; methodology, all; project administration, all; resources, all; software, all; supervision, all; validation, all; visualization, all; writing—original draft, all; writing—review & editing, all. All authors have read and agreed to the published version of this manuscript.

Funding: This work was partially supported by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, Project No. TKP2021-NVA-16 (I.I.R. and Z.B.) and TKP2021-NKTA-64 (Z.B.).

Data Availability Statement: The CRAN R code and data used for this study are publicly available at https://github.com/zbagoly/GRBPreconMap, accessed on 12 December 2025.

Acknowledgments: The authors thank Lajos G. Balázs and István Horváth for the enlightening discussions and the anonymous referees for valuable comments on this paper.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. The Advantage of Factorized Density Estimation in GRB Angular Maps

Estimating the spatial density of the GRBs with measured redshift is more efficient using the factorized approach—i.e., as the product of the Swift BAT sky exposure map and the redshift survey efficiency map—than using direct AKDE on the GRB subset. Traditional quadratic error metrics, such as Mean Integrated Square Error (MISE), encourage geometric smoothing, which is detrimental to preserving the structural fidelity of the redshift density map, especially at sharp geometric boundaries [38,41]. Here, we show that the bandwidth limitation of AKDE estimation can be derived by analyzing the Asymptotic Mean Square Error [37]. Using the language of Fourier signal processing, we show that the necessarily large bandwidth due to sparse sampling behaves as a strong low-pass filter, reaching the soft ceiling of resolution [51]. With Fourier transform we prove that the factorized estimator acts as a spectral bandwidth extender, recovering information that would be mathematically impossible to resolve based on the sparse sample of GRBs with redshift (z-GRBs) [52].

Appendix A.1. Fidelity Versus Smoothing

Let $\mathcal{S} \subset \mathbb{S}^2$ be the observation domain. We observe a realization of the inhomogeneous Poisson point process, $\mathcal{D}_{BAT}$, which contains N detections. In our case, this is the entire Swift BAT catalog. Within this, there is a marked subset $\mathcal{D}_z$ containing n GRBs with redshifts, where $n \ll N$ [20]. Let $\lambda_{BAT}(\mathbf{x})$ denote the intensity of the entire population, which is determined by the Swift BAT sky exposure map (the Modulator) [19]. Let $\epsilon(\mathbf{x})$ denote the spectroscopic redshift recovery efficiency (the Carrier Signal), which gives the conditional probability that a detected GRB has a measured redshift.

The true observational probability of GRBs with redshift, $\lambda_z(\mathbf{x})$, can be defined as follows:

$$\lambda_z(\mathbf{x}) = \lambda_{BAT}(\mathbf{x}) \cdot \epsilon(\mathbf{x}) \tag{A1}$$

We note that standard procedures aim to minimize MISE. However, for problems with strict geometric constraints (e.g., sharp visibility cutoff at the galactic disk), minimizing MISE is contraproductive, as it encourages significant geometric smoothing. This results in blurred boundaries that physically misrepresent the efficiency map [40]. Our goal is to maximize structural fidelity and final information content, and to choose between the Factorized Estimation and the AKDE method.

Under the Factorized Estimation, $\hat{\lambda}_z^{(Fact)} = \hat{\lambda}_{BAT} \times \hat{\epsilon}_{GAM}$ separates the BAT exposure from the redshift search efficiency, while using AKDE, it can be defined as $\hat{\lambda}_z^{(AKDE)} = \mathrm{AKDE}(\mathcal{D}_z)$.

Appendix A.2. The Bandwidth Limit and the Lower Pass Limit

To understand why AKDE cannot preserve the geometry of the survey, we need to examine how the bandwidth adapts to the sparse sample size n. In AKDE, the smoothing scale varies locally:

$$\hat{\lambda}_z^{(AKDE)}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h(\mathbf{x}_i)^2} K\left(\frac{\mathbf{x}-\mathbf{x}_i}{h(\mathbf{x}_i)}\right) \tag{A2}$$

where $h(\mathbf{x}_i)$ is the local bandwidth parameter assigned to the i-th data point, typically derived from a pilot density estimate $\tilde{\lambda}(\mathbf{x}_i)$ such that $h(\mathbf{x}_i) \propto \tilde{\lambda}(\mathbf{x}_i)^{-\alpha}$ (often $\alpha = 0.5$ according to Abramson's law) [36].

Although AKDE aims to reduce bias at the tails of distributions, in the case of variable bandwidth estimation, the local bandwidths $h(\mathbf{x}_i)$ are constrained by the pilot estimate, which itself relies on the sparse n samples. To ensure the stability of the pilot estimate (i.e., to control the variance term), the smoothing parameter h_0 must be scaled as follows [36,37]:

$$h_{opt} \propto n^{-1/(d+4)} = n^{-1/6} \quad \text{(in our case } d = 2\text{)} \tag{A3}$$

Even if the bandwidth is adaptive, the lower limit of the minimum bandwidth available in dense regions is determined by the global sample size n. Since n is small in the case of the redshift sample, the smoothing kernel functions remain relatively large, even in the densest regions. This large h acts as a low-pass filter, making it mathematically impossible to follow step-like occlusions that occurs, e.g., around the Galactic disk.

Appendix A.3. Signal Processing Analysis: Spectral Bandwidth Extension with Modulation/Demodulation

AKDE can be modeled as the convolution of empirical intensity and a variable kernel function $K_{h(\mathbf{x})}(\mathbf{x})$. In the frequency domain, this operation approximates multiplication by the average kernel width transfer function. For a Gaussian kernel, the transfer function is $\tilde{K}(\mathbf{k}) = e^{-\frac{1}{2}\bar{h}^2|\mathbf{k}|^2}$. This acts as a low-pass filter, where the effective cutoff frequency k_c is inversely proportional to the average bandwidth:

$$k_c \approx \frac{1}{\bar{h}} \propto n^{1/6} \tag{A4}$$

Since n is small, $\bar{h}$ is large, and, consequently, k_c is very low. Any spatial feature with a frequency $|\mathbf{k}| > k_c$—such as sharp optical occlusions, e.g., at the Galactic disk region—is attenuated to near zero. AKDE inevitably blurs the boundaries, leaking probability mass into unobserved regions (aliasing), and encounters a soft ceiling on resolution [51].

Let $\mathcal{F}\{f\} = \tilde{\Lambda}(\mathbf{k})$ be the Fourier transform of F. Here, intensity is modeled as a modulation process, where the carrier is the efficiency $\epsilon(\mathbf{x})$, which represents the redshift measurement. This could contains step-like terms, which act as a sharp cutoff that is difficult to model. This sharp cutoff shows a wideband signal with a large bandwidth (e.g., $|\mathbf{k}|^{-1}$). The Modulator function, λ_{BAT}, represents the Swift BAT sky exposure map. It is typically smooth and changes slowly across the sky; so, this is our narrowband signal.

The factorized estimate constructs the spectrum in Fourier space using convolution:

$$\tilde{\Lambda}_z^{(Fact)}(\mathbf{k}) = (\tilde{\Lambda}_{BAT} * \tilde{E}_{GAM})(\mathbf{k}) \tag{A5}$$

Here, we first estimate the λ_{BAT} exposure from the complete N-element GRB catalog, which resolves the fundamental modulator function. The estimated ϵ efficiency is determined using Generalized Additive Models (GAMs) from the entire dataset, based on N events

(binary output). The convolution mixes the high-frequency components of the efficiency map back into the following result:

$$\text{Support}(\tilde{\Lambda}_B^{(Fact)}) \approx \text{Support}(\tilde{\Lambda}_T) + \text{Support}(\tilde{P}) \approx \text{Wideband} \tag{A6}$$

Since $N \gg n$, we are able to capture the high frequencies of the carrier signal ($\Omega_{eff} \gg k_c$), including, e.g., sharp changes in the map. The factorized estimate effectively performs spectral bandwidth expansion, using the entire dataset to recover spatial features that are not accessible by the direct AKDE method.

Appendix A.4. Information Theory Approach

To further quantify the strength of the factorized estimator, we apply Shannon's information theory framework, where density estimation is treated as signal reconstruction from noisy channel measurements. The signal is the true intensity field $\lambda_z(\mathbf{x})$, and the noise arises from the stochastic nature of the Poisson point process. The fundamental measure of statistical discrepancy between the true intensity $\lambda(\mathbf{x})$ and the estimator $\hat{\lambda}(\mathbf{x})$ is the generalized Kullback–Leibler (KL) divergence (or relative entropy):

$$D_{KL}(\lambda||\hat{\lambda}) = \int_{\mathcal{S}} \left(\lambda(\mathbf{x}) \ln \frac{\lambda(\mathbf{x})}{\hat{\lambda}(\mathbf{x})} - \lambda(\mathbf{x}) + \hat{\lambda}(\mathbf{x}) \right) d\mathbf{x} \tag{A7}$$

Minimizing the divergence is asymptotically equivalent to maximizing the likelihood of the observed data.

In the factorized approach, we separate the two true intensity factors: $\lambda_z(\mathbf{x}) = \lambda_{BAT}(\mathbf{x}) \cdot \epsilon(\mathbf{x})$. Substituting the estimator $\hat{\lambda}_z^{(Fact)} = \hat{\lambda}_{BAT}\hat{\epsilon}$ into the KL definition and assuming that the estimators are reasonably close to reality (meaning that the second-order cross-term is negligible), the total information loss decomposes additively:

$$D_{KL}(\lambda_z||\hat{\lambda}_z^{(Fact)}) \approx \underbrace{D_{KL}(\epsilon||\hat{\epsilon})}_{\text{Redshift Effect Estimation Error}} + \underbrace{\int_{\mathcal{S}} \epsilon(\mathbf{x}) D_{KL}(\lambda_{BAT}(\mathbf{x})||\hat{\lambda}_{BAT}(\mathbf{x}))d\mathbf{x}}_{\text{BAT Exposure Map Reconstruction Error}} \tag{A8}$$

This decomposition proves that the total error can be minimized by independently minimizing the efficiency function (ϵ) and the exposure map reconstruction (λ_{BAT}) errors. The redshift detection efficiency error is minimized by the GAM using the entire N dataset, which captures the high-frequency components, while the BAT exposure map error is minimized by using the high-bandwidth $\mathcal{D}_{BAT}$ dataset (N detections).

By using narrow bandwidth, AKDE irreversibly destroys the fine scale spatial information. The loss of information regarding the BAT background events (i.e., all the events), is also manifested spectrally: discarding the high-density background reduces the effective sampling rate below the threshold required for fine resolution. Consequently, reconstructing this high-bandwidth signal from sparse samples violates the Nyquist–Shannon sampling theorem, resulting in aliasing and information loss. In contrast, the factorized estimate preserves this information by using $\mathcal{D}_{BAT}$ to construct the carrier density, thus operating under a strictly higher mutual information constraint than the direct estimator.

From the Maximum Entropy perspective, this means that results must be achieved in accordance with the distribution, maximizing entropy under the constraints provided by the data. AKDE is limited only by n redshift points; so, in regions of the redshift efficiency map with near-zero visibility, it spreads the mass evenly due to smoothing, creating false signals or leaks. The factorized estimate is limited by N complete detections, which acts as a hard prior and forces $\hat{\epsilon}(\mathbf{x}) \approx 0$ in the low-probability areas. Since $\hat{\lambda}_z = \hat{\lambda}_{BAT} \times \hat{\epsilon}$, this approaches zero for the final map density, preserving the physical reality of the survey

constraints. The factorized estimate effectively performs Conditional Density Estimation, injecting structural constraints from the survey that are missing from the sparse subset of AKDE, which strictly reduces the Kullback–Leibler divergence between the estimate and reality.

AKDE is forced to smooth the data ($h \propto n^{-1/6}$) in order to control variance, thereby creating a low-pass filter that destroys the high-frequency information in the redshift survey efficiency map. The factorized estimator avoids this by separating the large-scale background (Carrier) from the intrinsic Swift BAT sky exposure map (Modulator). This works as a signal recovery technique, using the entire BAT catalog extra positional information to overcome the bandwidth limitation imposed by the sparse redshift sample.

Appendix B. Methodology: The Spherical Generalized Additive Model

In order to avoid the issues of the KDE ratio, we reformulate the task not as a density ratio but as a conditional probability estimation problem (binary classification). For each i-th event ($i = 1 \dots N_{total}$), we assign an indicator variable $y_i \in \{0, 1\}$:

$$y_i = \begin{cases} 1 & \text{if the event } \in \mathcal{D}_z \text{ (redshift determined)} \\ 0 & \text{if the event } \in \mathcal{D}_{total} \setminus \mathcal{D}_z \text{ (no redshift determined)} \end{cases} \tag{A9}$$

The goal is to estimate the posterior probability $p(\mathbf{x}) = P(Y = 1|\mathbf{x})$, where $\mathbf{x} \in \mathbb{S}^2$ denotes the location vector.

Here, we choose to estimate the conditional probability with a Generalized Additive Model [53,54]. Since the target variable is binary ($y_i \in \{0, 1\}$), the probability variable is described by a Bernoulli distribution, denoted by $Y \sim \text{Bernoulli}(p(\mathbf{x}))$. This is the most economical discrete distribution, which is completely characterized by a single parameter, the p success probability (i.e., the probability of determining redshift). Within the GAM framework, the linear predictor (η) and the expected value ($E[Y] = p$) are linked by a monotonic, differentiable link function ($g(\cdot)$) such that $\eta = g(p)$. In this case, we use the canonical link function, the logit link function:

$$\eta(\mathbf{x}) = \text{logit}(p(\mathbf{x})) = \ln\left(\frac{p(\mathbf{x})}{1 - p(\mathbf{x})}\right) = f(\mathbf{x}) \tag{A10}$$

where $f(\mathbf{x})$ is a smooth function defined on the spherical manifold.

The primary motivation in applying the logit transformation is to remove range constraints: while the $f(\mathbf{x})$ linear predictor can take values on the entire real number line ($\mathbb{R}$, i.e., $(-\infty, +\infty)$), the $p(\mathbf{x})$ probability is, by definition, strictly limited to the unit interval $[0, 1]$.

The logit function, as the inverse of the logistic (or sigmoid) function, establishes a bijective mapping between these two domains, ensuring that the probabilities estimated by the model remain physically valid (strictly within the limits), a condition that would not be met when using simple linear regression.

The Bernoulli distribution is characterized by a natural parameter, which is $\ln(p/(1 - p))$. If the chosen link function is equal to the natural parameter, it is called the canonical link. The use of the canonical link automatically ensures the positive definiteness of the Fisher information matrix, which, in turn, guarantees the concavity of the log-likelihood function. It guarantees that during numerical optimization (e.g., Newton–Raphson or P-IRLS), the solution converges to a single global optimum, eliminating the risk of becoming stuck in local maxima.

Logistic regression is also characterized by information-theoretic consistency: it can be shown that this model represents the Maximum Entropy (MaxEnt) model under the constraint that the expected values of the predictors in the model match the empirical

means [55]. This model makes the fewest hidden assumptions about data distribution, providing the least biased estimate based on the existing information.

The logit model can be interpreted in the context of Bayes' theorem [56]. Consider the posterior probability $P(\text{redshift}|\mathbf{x})$, which represents the probability that an event observed at position $\mathbf{x}$ has a measured redshift. According to Bayes' theorem, it is the product of the likelihood ($P(\mathbf{x}|\text{redshift})$) and the prior ($P(\text{redshift})$).

This method effectively leads to the development of a discriminative model. Instead of trying to estimate the generative density functions ($P(\mathbf{x}|\text{redshift})$ and $P(\mathbf{x}|\text{no redshift})$) independently (which is the basis of the KDE ratio method), we learn the decision boundary and the posterior probability surface directly using logistic regression. This is advantageous because estimating generative densities is a significantly more difficult task, especially with sparse data or in high-dimensional spaces (due to the 'curse of dimensionality'), while discriminative modeling focuses on separating classes, which is often robustly achievable even with smaller sample sizes [57].

In astrophysics, this methodology facilitates the precise handling of 'contamination' and 'completeness'. If we visualize the probability $P(\text{redshift}|\mathbf{x})$ as a map, then the regions where the relative density of redshift determinations is significantly higher than the background are directly highlighted, regardless of the absolute density fluctuations in the entire sample (e.g., the effects of galactic extinction).

Appendix B.1. Spline on the Sphere

The central element of the model is the estimation of the angular function $f(\mathbf{x})$. Traditional methods operating in the Euclidean plane, such as the application of tensor-product B-splines [58] to spherical coordinates (θ, ϕ), cause serious topological anomalies. The most significant of these is the pole singularity, since the length of the latitudes decreases to zero at the poles. Consequently, grid-based smoothing results in artifacts, causing artificial 'knotting' at the poles. Similarly problematic is the issue of cyclicity, since continuity and differentiability must be enforced at the boundaries $-180°$ and $+180°$ for the longitude (ϕ) (periodic boundary condition). To solve these problems, we apply the spherical extension of Thin Plate Splines (TPS) introduced by Duchon [59], known as the Splines on the Sphere method, which was further developed for statistical applications by Wahba [60] and Wood [61].

The basic assumption of TPS is to find a function $f(\mathbf{x})$ that minimizes the penalized least squares error function, where the penalty term quantifies the surface curvature (i.e., 'the bending energy'). In the Euclidean plane, this penalty term is defined as the integral of the quadratic second derivatives, where Duchon [59] solved this variational problem by using a sum of radial basis functions. The functional form of the basis function is determined by the dimension of the space and the order of the derivatives; in 2D, this corresponds to the well-known $r^2 \cdot \ln(r)$. However, the sphere ($\mathbb{S}^2$) is not a Euclidean space; so, the direct application of planar kernel functions would lead to distortions due to metric differences in distance measurement. Wahba [60] extended Duchon's theory to the sphere, defining the corresponding smoothness functional via the Laplace–Beltrami operator. The spherical bending energy is given by $J(f) = \int_{\mathbb{S}^2} (\Delta f)^2 d\sigma$, where Δ is the spherical Laplace operator. Minimizing this functional yields spherical Green functions. Wahba derived the basis function (core function) of spherical splines in the following closed-form expression:

$$\psi(r) = \frac{1}{4\pi}\left[\ln\left(\sin\left(\frac{r}{2}\right)\right) + \frac{1}{2} - \frac{1}{2}\ln 2\right] \tag{A11}$$

where r denotes the geodesic angular distance between the two points (along the main circle) in radians. This function exhibits a singularity at $r = 0$ (it tends to $-\infty$ due to the logarithmic term), which is why it is called pseudo-singular. Although the basis

functions themselves are singular, Wood [61] proved that within the statistical model, the total estimated surface remains globally continuous and finite. This can be attributed to constraints on the model coefficients (δ_j) (specifically, perpendicularity to the constant term), which ensure that singularities are cancelled out. In practice, we construct the model in the form $f(\mathbf{x}) = \beta_0 + \sum_{j=1}^{k} \delta_j \psi(\text{dist}(\mathbf{x}, \mathbf{x}_j))$, where $\mathbf{x}_j$ forms a set of uniformly distributed nodes on the sphere. This representation is extremely advantageous because the basis functions are isotropic (they depend only on distance); so, the model is completely invariant with respect to the orientation of the coordinate system; there are no privileged directions or poles.

For computational efficiency, we avoid using all data points as nodes (which would require $N \times N$ matrix inversion). Instead, we use a low-rank approximation. Wood [61] introduced a procedure based on eigenvalue decomposition for projecting the entire basis function set onto a lower-dimensional ($k \ll N$) subspace. This projection is optimal in the sense that the approximation error measured in the Frobenius norm is minimal. Thus, the method is capable of handling datasets containing tens of thousands of elements (such as the value N_{total}) while preserving mathematical precision and topological correctness. The SOS method provides the smoothest possible interpolation on the sphere defined by the surface curvature that fits the data. The degree of smoothness is modulated by the λ parameter in the penalized likelihood, which penalizes excessive surface curvature. This approach eliminates the fitting problems experienced with local polynomial fitting (e.g., LOESS) and the subjectivity associated with bandwidth selection, as smoothness is determined globally by the model based on variational theory.

The preference for the SOS continuous spline model over the grid-discretized HEALPix framework is driven by the fundamental limitations of fixed-resolution pixelization in density estimation. While HEALPix facilitates fast indexing, its globally fixed resolution (N_{SIDE}) necessitates a compromise: low resolutions obscure fine details (underfitting), whereas high resolutions in sparsely populated regions introduce significant stochastic shot noise. In contrast, the SOS method employs an adaptive, continuous function space that enables sub-pixel interpolation, resolving detailed structures where data is abundant while naturally smoothing sparse regions. Although formulating a General Additive Model on a HEALPix basis is mathematically feasible, it yields a computationally prohibitive, high-dimensional sparse design matrix ($N_{\text{pix}} \approx 10^5 - 10^7$). The SOS implementation by Wood [61] circumvents this optimization bottleneck by utilizing a low-rank basis of optimal eigenfunctions, drastically reducing computational complexity without compromising expressive power, thereby offering a more statistically efficient and objective Bayesian framework for reconstructing the underlying latent probability field.

Appendix B.2. Bayesian Estimation and the Duality of Penalized Likelihood

Parameter estimation ($\boldsymbol{\beta}$) cannot rely on the traditional Maximum Likelihood (ML) method. The reason for this is that ML, especially when using a large-dimensional spline basis, is highly prone to overfitting, leading to excessive variance in the estimated function [62–64]. As a result, the model would follow the measurement noise and stochastic fluctuations rather than the underlying smooth signal, causing the surface to show excessive 'waviness'. For robust estimation, we therefore use penalized likelihood estimation [65,66]:

$$\mathcal{L}_p(\boldsymbol{\beta}) = \mathcal{L}(\boldsymbol{\beta}) - \frac{1}{2} \sum_i \lambda_i \boldsymbol{\beta}^T \mathbf{S}_i \boldsymbol{\beta} \tag{A12}$$

where $\mathcal{L}$ denotes the binomial log-likelihood, $\mathbf{S}_i$ is the smoothing penalty matrix, and λ_i is the smoothing parameter.

The penalty term ($\sum_i \boldsymbol{\beta}^T \mathbf{S}_i \boldsymbol{\beta}$) quantifies the bending energy of the surface, which is the discretized form of the integral of the second derivatives (curvature) over the spherical surface:

$$J(f) = \int_{\mathbb{S}^2} (\Delta f)^2 d\sigma \tag{A13}$$

The balance between data fidelity (fit) and smoothness (regularization) is controlled by the smoothing parameter λ.

Appendix B.3. Regularization and the Role of Gaussian Priors

The penalized likelihood method is based on regularization theory [65], which limits model complexity in a mathematically equivalent way to Tikhonov regularization, using a smoothness penalty term. In Bayesian terms, the optimization problem (maximizing $\mathcal{L}_p(\boldsymbol{\beta})$) strictly corresponds to the Maximum A Posteriori (MAP) estimate, assuming that the $\boldsymbol{\beta}$ parameters follow a central multivariate Gaussian prior distribution [67]:

$$\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, \mathbf{S}_\lambda^-), \quad \text{where} \quad \mathbf{S}_\lambda^- \propto \frac{1}{\lambda} \mathbf{S}^{-1}. \tag{A14}$$

This prior physically expresses the smoothness of the objective function, $f(\mathbf{x})$, where the parameter λ controls the precision of the prior. The limiting case $\lambda \to \infty$ leads to zero variance and an 'infinitely smooth' surface, while $\lambda \to 0$ results in the model reverting to traditional Maximum Likelihood estimation. The λ hyperparameter is a key element of the model, which directly controls the bias-variance trade-off: insufficient regularization leads to noise modeling (overfitting), while excessive penalty leads to loss of structural information (underfitting). Here, λ is not an arbitrary choice but a hyperparameter that can be estimated objectively from the data, ensuring reliability even in the case of noisy astrophysical data [68].

Appendix B.4. The Restricted Maximum Likelihood Method

For objective estimation of λ, we use the Restricted Maximum Likelihood method, which was originally introduced for variance component estimation [69] and subsequently extended to generalized linear models by Wood [66].

REML is efficient because it estimates the λ hyperparameter by maximizing a restricted likelihood function, making it independent of the point estimates of fixed effects. This approach corrects for the bias of the traditional ML method on finite samples, as it explicitly accounts for the degrees of freedom lost in the estimation of the $\boldsymbol{\beta}$ parameters. REML is equal to maximizing the Bayesian evidence, where the $\boldsymbol{\beta}$ coefficients are treated as nuisance parameters and are integrated out.

The structure of the posterior distribution can be written as follows using Bayes' theorem:

$$P(\boldsymbol{\beta}, \lambda | \mathbf{y}) \propto P(\mathbf{y} | \boldsymbol{\beta}, \lambda) P(\boldsymbol{\beta} | \lambda) P(\lambda) \tag{A15}$$

Here, the goal is to maximize the marginal posterior distribution ($P(\lambda|\mathbf{y})$) of the hyperparameter λ, which is obtained by integrating over the entire posterior $\boldsymbol{\beta}$:

$$P(\lambda | \mathbf{y}) = \int P(\boldsymbol{\beta}, \lambda | \mathbf{y}) d\boldsymbol{\beta} \propto P(\lambda) \int P(\mathbf{y} | \boldsymbol{\beta}, \lambda) P(\boldsymbol{\beta} | \lambda) d\boldsymbol{\beta} \tag{A16}$$

Assuming a non-informative (flat) $P(\lambda)$ prior, the logarithm of the marginal likelihood takes the following form:

$$\mathcal{V}(\lambda) = \log \int \exp\left(\mathcal{L}(\boldsymbol{\beta}) - \frac{1}{2} \boldsymbol{\beta}^T \mathbf{S}_\lambda \boldsymbol{\beta} \right) d\boldsymbol{\beta} \tag{A17}$$

where $\mathcal{L}(\boldsymbol{\beta})$ is the log-likelihood, and the second term is the penalty term derived from the Gaussian prior. Maximizing λ thus selects the level of smoothness that best explains the data when averaged over the entire parameter space (rather than at a single optimal point). This mechanism, through the Occam factor, naturally penalizes unnecessary complexity, since in overly complex models, a significant portion of the probability mass is distributed in regions that are incompatible with the data.

Since this integral cannot be solved analytically in general, we use the Laplace approximation in practice. The Laplace method approximates the integrand with a multivariate Gaussian distribution around the mode (the MAP estimate $\hat{\boldsymbol{\beta}}$). Therefore, the REML criterion takes (approximately) the following form:

$$\mathcal{V}_{REML}(\lambda) \approx \mathcal{L}_p(\hat{\boldsymbol{\beta}}) - \frac{1}{2}\log|\mathbf{H}_\lambda| + \frac{M}{2}\log(2\pi) \tag{A18}$$

where $\mathbf{H}_\lambda$ is the Hessian matrix of the penalized likelihood function (the matrix of second derivatives) at point $\hat{\boldsymbol{\beta}}$, which can be defined as the sum of the Fisher information matrix and the penalty matrix: $\mathbf{H}_\lambda = \mathcal{I}(\hat{\boldsymbol{\beta}}) + \lambda\mathbf{S}$. The term $\log|\mathbf{H}_\lambda|$ (the logarithm of the determinant of the Hessian matrix) penalizes model complexity and accounts for the degrees of freedom correction.

By correcting for the uncertainty arising from the estimation of $\boldsymbol{\beta}$, the REML method reduces bias, resulting in more accurate variance estimates, especially for limited sample sizes. In the context of spherical data, where the assumption of independence of errors is often violated, REML exhibits excellent robustness to angular autocorrelation. It is less prone to over-smoothing than Generalized Cross Validation or AIC-based smoothing selection, which tend to misinterpret noise as signal in the presence of correlated errors [70]. Finally, the method ensures Bayesian consistency, as it closely fits the 'Empirical Bayesian' or 'Type II Maximum Likelihood' approach, where the parameters of the prior distribution (in this case, the degree of smoothness) are learned from the data, avoiding subjective manual tuning.

From an information theory perspective, the REML method is related fully to the principle of entropy maximization. Shannon entropy measures the uncertainty of information, and according to the Maximum Entropy principle, the most appropriate model is the one that makes the fewest assumptions about unknown data while satisfying the observation constraints [52]. In the REML framework, the use of Gaussian priors on the parameters corresponds to such a Maximum Entropy assumption with fixed first and second moments (expectation and covariance). Maximizing the marginal likelihood (or Bayesian evidence) is equivalent to optimizing the information content represented by the model, where the penalty term (the logarithm of the determinant of the Hesse matrix) enforces Occam's razor, favoring simpler (smoother) models with lower information content if they explain the data with equal efficiency. Caticha [71] showed that Bayes' rule can also be derived from purely entropy-based reasoning using the method of entropic inference. In this context, hyperparameter optimization performed by REML can be viewed as minimizing the relative entropy (Kullback–Leibler divergence) between the true, unknown data-generating process and the distributions covered by the model family.

Appendix B.5. Model Selection and AIC

While the smoothing parameter λ is optimized using the REML method, the base dimension (k) is a structural hyperparameter that limits the maximum theoretical complexity of the model. Although the penalty term effectively controls the effective degrees of freedom (EDF), choosing an inappropriate value for k carries risks: too low a value leads to underfitting (bias), while too high a dimension places an unnecessary computational

burden on the analysis. To determine the optimal value of k, we use the Akaike Information Criterion, which quantifies the trade-off between model fit and complexity [72]:

$$\mathrm{AIC} = -2\ln(\hat{L}) + 2 \cdot \mathrm{EDF} \tag{A19}$$

where $\hat{L}$ is the maximized likelihood value and EDF is the effective degrees of freedom defined by the trace of the penalty matrix.

The theoretical basis of AIC is the Kullback–Leibler (KL) divergence, which measures the information loss between two probability distributions. The goal of AIC is to asymptotically estimate the relative expected KL divergence between the distribution generated by the model and the unknown 'true' data-generating mechanism. Selecting the minimum AIC corresponds to the model leading to minimum information loss relative to reality. The 2EDF penalty term in the criterion corrects for the bias that arises from fitting the model and evaluating its performance on the same dataset. The k parameter is optimized using grid search, selecting the configuration that minimizes the AIC value globally, thereby validating the principle of parsimony (economy).

We must emphasize that the AIC does not have an absolute scale; only the relative difference between models ($\Delta AIC_i = AIC_i - AIC_{min}$) carries relevant information for model selection. Based on the widely accepted empirical guidelines of Burnham and Anderson [49], if ΔAIC falls between 0 and 2, the models examined can be considered statistically indistinguishable; so, the more complex model does not offer a significant advantage. A difference between 4 and 7 provides moderate evidence against the model with the higher AIC value (weaker performance), indicating that the fit of the model with the lower value is noticeably better. A ΔAIC value exceeding 10 is considered strong evidence against the weaker model, effectively ruling out the possibility that it is the optimal choice from the given set of models for describing the data under investigation.

Appendix B.6. R Implementation

The above procedure was implemented in the R programming environment using the *mgcv* package. Our model fitting uses GAMs, which are based on the Penalized Iterative Reweighted Least Squares (P-IRLS) algorithm optimized for non-Gaussian probabilities. The smoothing parameter (λ) was determined by using REML estimation, eliminating bias, while the optimal complexity of the model (the k basis dimension) was selected by finding the global minimum of the AIC. With this optimization process, the final high-resolution grid-predicted probability map provides an unbiased and economical estimate that fits the internal structure of the data. The used code and data are publicly available at https://github.com/zbagoly/GRBPreconMap, accessed on 12 December 2025.

Notes

1 https://swift.gsfc.nasa.gov/archive/grb_table/
2 http://www.swift.ac.uk/xrt_live_cat/
3 https://www.mpe.mpg.de/~jcg/grbgen.html
4 https://sites.astro.caltech.edu/grbox/grbox.php

References

1. Feigelson, E.D.; Babu, G.J. *Modern Statistical Methods for Astronomy*; Cambridge University Press: Cambridge, UK, 2012.
2. Mészáros, P. Gamma-ray bursts. *Rep. Prog. Phys.* **2006**, *69*, 2259–2321. [CrossRef]
3. Mészáros, A.; Bagoly, Z.; Balázs, L.G.; Horváth, I. Redshift distribution of gamma-ray bursts and star formation rate. *Astron. Astrophys.* **2006**, *455*, 785–790. [CrossRef]
4. Tugay, A.; Tarnopolski, M. Continuous Filament Network of the Local Universe. *Astrophys. J.* **2023**, *952*, 3. [CrossRef]

5. Wang, H.; Liang, N. Constraints from Fermi observations of long gamma-ray bursts on cosmological parameters. *Mon. Not. R. Astron. Soc.* **2024**, *533*, 743–755. [CrossRef]
6. Greiner, J.; Iyudin, A.; Kanbach, G.; Zoglauer, A.; Diehl, R.; Ryde, F.; Hartmann, D.; Kienlin, A.V.; McBreen, S.; Ajello, M.; et al. Gamma-ray burst investigation via polarimetry and spectroscopy (GRIPS). *Exp. Astron.* **2009**, *23*, 91–120. [CrossRef]
7. Liu, T.; Kim, K.T.; Yoo, H.; Liu, S.y.; Tatematsu, K.; Qin, S.L.; Zhang, Q.; Wu, Y.; Wang, K.; Goldsmith, P.F.; et al. Star Formation Laws in Both Galactic Massive Clumps and External Galaxies: Extensive Study with Dust Coninuum, HCN (4-3), and CS (7-6). *Astrophys. J.* **2016**, *829*, 59. [CrossRef]
8. Dainotti, M.G.; Petrosian, V.; Bowden, L. Cosmological Evolution of the Formation Rate of Short Gamma-Ray Bursts with and without Extended Emission. *Astrophys. J.* **2021**, *914*, L40. [CrossRef]
9. Dong, X.F.; Li, X.J.; Zhang, Z.B.; Zhang, X.L. A comparative study of luminosity functions and event rate densities of long GRBs with non-parametric method. *Mon. Not. R. Astron. Soc.* **2022**, *513*, 1078–1087. [CrossRef]
10. Dong, X.F.; Huang, Y.F.; Zhang, Z.B.; Geng, J.J.; Deng, C.; Zou, Z.C.; Hu, C.R.; Amat, O. A Statistical Study of the Gamma-Ray Burst and Supernova Association. *Astrophys. J.* **2025**, *993*, 20. [CrossRef]
11. Yadav, J.K.; Bagla, J.S.; Khandai, N. Fractal dimension as a measure of the scale of homogeneity. *Mon. Not. R. Astron. Soc.* **2010**, *405*, 2009–2015. [CrossRef]
12. Balázs, L.G.; Bagoly, Z.; Hakkila, J.E.; Horváth, I.; Kóbori, J.; Rácz, I.I.; Tóth, L.V. A giant ring-like structure at $0.78 < z < 0.86$ displayed by GRBs. *Mon. Not. R. Astron. Soc.* **2015**, *452*, 2236–2246. [CrossRef]
13. Balázs, L.G.; Rejtó, L.; Tusnády, G. Some statistical remarks on the giant GRB ring. *Mon. Not. R. Astron. Soc.* **2018**, *473*, 3169–3179. [CrossRef]
14. Horváth, I.; Hakkila, J.; Bagoly, Z. Possible structure in the GRB sky distribution at redshift two. *Astron. Astrophys.* **2014**, *561*, L12. [CrossRef]
15. Horváth, I.; Bagoly, Z.; Hakkila, J.; Tóth, L.V. New data support the existence of the Hercules-Corona Borealis Great Wall. *Astron. Astrophys.* **2015**, *584*, A48. [CrossRef]
16. Horvath, I.; Szécsi, D.; Hakkila, J.; Szabó, Á.; Racz, I.I.; Tóth, L.V.; Pinter, S.; Bagoly, Z. The clustering of gamma-ray bursts in the Hercules-Corona Borealis Great Wall: The largest structure in the Universe? *Mon. Not. R. Astron. Soc.* **2020**, *498*, 2544–2553. [CrossRef]
17. Horvath, I.; Bagoly, Z.; Balazs, L.G.; Hakkila, J.; Horvath, Z.; Joo, A.P.; Pinter, S.; Tóth, L.V.; Veres, P.; Racz, I.I. Mapping the Universe with gamma-ray bursts. *Mon. Not. R. Astron. Soc.* **2024**, *527*, 7191–7202. [CrossRef]
18. Horvath, I.; Bagoly, Z.; Balazs, L.G.; Hakkila, J.; Koncz, B.; Racz, I.I.; Veres, P.; Pinter, S. Scanning the Universe for Large-Scale Structures Using Gamma-Ray Bursts. *Universe* **2025**, *11*, 121. [CrossRef]
19. Baumgartner, W.H.; Tueller, J.; Markwardt, C.B.; Skinner, G.K.; Barthelmy, S.; Mushotzky, R.F.; Evans, P.A.; Gehrels, N. The 70-Month Swift-BAT All-sky Hard X-ray Survey. *Astrophys. J. Suppl. Ser.* **2013**, *207*, 19. [CrossRef]
20. Sakamoto, T.; Barthelmy, S.D.; Barbier, L.; Cummings, J.R.; Fenimore, E.E.; Gehrels, N.; Hullinger, D.; Krimm, H.A.; Markwardt, C.B.; Palmer, D.M.; et al. The First Swift BAT Gamma-Ray Burst Catalog. *Astrophys. J.* **2008**, *175*, 179–190. [CrossRef]
21. Okabe, A.; Boots, B.; Sugihara, K.; Chiu, S.N. Spatial Interpolation. In *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2008.
22. Schaap, W.E.; van de Weygaert, R. Continuous Fields and Discrete Samples: Reconstruction through Delaunay Tessellations. *Astron. Astrophys.* **2000**, *363*, L29–L32.
23. Garcia-Quintero, C.; Noriega, H.E.; de Mattia, A.; Aviles, A.; Lodha, K.; Chebat, D.; Rohlf, J.; Nadathur, S.; Elbers, W.; Aguilar, J.; et al. Cosmological implications of DESI DR2 BAO measurements in light of the latest ACT DR6 CMB data. *Phys. Rev. D* **2025**, *112*, 083529. [CrossRef]
24. Euclid Collaboration; Scaramella, R.; Amiaux, J.; Mellier, Y.; Burigana, C.; Carvalho, C.S.; Cuillandre, J.C.; Da Silva, A.; Derosa, A.; Dinis, J.; et al. Euclid preparation. I. The Euclid Wide Survey. *Astron. Astrophys.* **2022**, *662*, A112. [CrossRef]
25. Bagoly, Z.; Balázs, L.G.; Horváth, I.; Rácz, I.; Tóth, L.V.; Hakkila, J. The GRB's Sky Exposure Function. In Proceedings of the PoS (SWIFT 10)060, Swift: 10 Years of Discovery, Proceedings of Science, Rome, Italy, 2–5 December 2014. [CrossRef]
26. Bernardini, M.G.; Cordier, B.; Wei, J. The SVOM Mission. *Galaxies* **2021**, *9*, 113. [CrossRef]
27. Yuan, W.; Zhang, C.; Chen, C.; Ling, Z.; Lu, Y.; Osborne, J.P.; O'Brien, P.; Willingale, R.; Guainazzi, M.; Sodari, S.; et al. Science objectives of the Einstein Probe mission. *Sci. China Phys. Mech. Astron.* **2025**, *68*, 229502. [CrossRef]
28. Amati, L.; O'Brien, P.; Götz, D.; Boella, G.; Schanne, S.; Vermeulen, G.; Blain, A.; Tanvir, N.; Ghirlanda, G.; Castro-Tirado, A.J.; et al. The THESEUS space mission concept: Science case, design and expected performances. *Adv. Space Res.* **2018**, *62*, 191–244. [CrossRef]
29. Rácz, I.I.; Hortobagyi, A.J. Studying the variability of the X-ray spectral parameters of high-redshift GRBs' afterglows. *Astron. Nachrichten* **2018**, *339*, 347–351. [CrossRef]
30. Barthelmy, S.D.; Barbier, L.M.; Cummings, J.R.; Fenimore, E. The Burst Alert Telescope (BAT) on the SWIFT Midex Mission. *Space Sci. Rev.* **2005**, *120*, 143–164. [CrossRef]

31. Gehrels, N.; Chincarini, G.; Giommi, P.; Mason, K.O.; Nousek, J.A.; Wells, A.A.; White, N.E.; Barthelmy, S.D.; Burrows, D.N.; Cominsky, L.R.; et al. The Swift Gamma-Ray Burst Mission. *Astrophys. J.* **2004**, *611*, 1005–1020. [CrossRef]
32. Goad, M.R.; Tyler, L.G.; Beardmore, A.P.; PEvans, A.; Rosen, S.R.; Osborne, J.P.; Starling, R.L.C.; Marshall, F.E.; Yershov, V.; Burrows, N.D.; et al. Accurate early positions for Swift GRBs: Enhancing X-ray positions with UVOT astrometry. *Astron. Astrophys.* **2007**, *476*, 1401–1409. [CrossRef]
33. Evans, P.A.; Beardmore, A.P.; Page, K.L.; Osborne, J.P.; O'Brien, P.T.; Willingale, R.; Starling, R.L.C.; Burrows, D.N.; Godet, O.; Vetere, L.; et al. Methods and results of an automatic analysis of a complete sample of Swift-XRT observations of GRBs. *Mon. Not. R. Astron. Soc.* **2009**, *397*, 1177–1201. [CrossRef]
34. Horvath, I.; Racz, I.I.; Bagoly, Z.; Balázs, L.G.; Pinter, S. Does the GRB Duration Depend on Redshift? *Universe* **2022**, *8*, 221. [CrossRef]
35. Bagoly, Z.; Horvath, I.; Racz, I.I.; Balázs, L.G.; Tóth, L.V. The Spatial Distribution of Gamma-Ray Bursts with Measured Redshifts from 24 Years of Observation. *Universe* **2022**, *8*, 342. [CrossRef]
36. Abramson, I.S. On Bandwidth Variation in Kernel Estimates—A Square Root Law. *Ann. Stat.* **1982**, *10*, 1217–1223. [CrossRef]
37. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; Chapman and Hall: London, UK, 1986. [CrossRef]
38. Ruppert, D.; Wand, M.P.; Carroll, R.J. *Semiparametric Regression*; Cambridge Series in Statistical and Probabilistic Mathematics; Cambridge University Press: Cambridge, UK, 2003.
39. Kerscher, M.; Szapudi, I.; Szalay, A.S. A comparison of estimators for the two-point correlation function. *Astrophys. J.* **2000**, *535*, L13. [CrossRef]
40. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2009. [CrossRef]
41. Shimazaki, H.; Shinomoto, S. Kernel bandwidth optimization in spike rate estimation. *J. Comput. Neurosci.* **2010**, *29*, 171–182. [CrossRef] [PubMed]
42. Hall, P.; Watson, G.S.; Cabrera, J. Kernel density estimation with spherical data. *Biometrika* **1987**, *74*, 751–762. [CrossRef]
43. Fisher, R.A. Dispersion on a sphere. *Proc. R. Soc. Lond. Ser. Math. Phys. Sci.* **1953**, *217*, 295–305. [CrossRef]
44. Mardia, K.V.; Jupp, P.E. *Directional Statistics*; John Wiley & Sons: Chichester, UK, 2008.
45. Stone, M. Cross-Validatory Choice and Assessment of Statistical Predictions. *J. R. Stat. Soc. Ser. B* **1974**, *36*, 111–133. [CrossRef]
46. Kass, R.E.; Raftery, A.E. Bayes factors. *J. Am. Stat. Assoc.* **1995**, *90*, 773–795. [CrossRef]
47. Phillips, S.J.; Anderson, R.P.; Schapire, R.E. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* **2006**, *190*, 231–259. [CrossRef]
48. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [CrossRef]
49. Burnham, K.P.; Anderson, D.R. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed.; Springer: New York, NY, USA, 2002.
50. Pinter, S.; Balazs, L.G.; Bagoly, Z.; Toth, L.V.; Racz, I.I.; Horvath, I. Some statistical remarks on GRBs jointly detected by Fermi and Swift satellites. *Mon. Not. R. Astron. Soc.* **2024**, *527*, 8931–8940. [CrossRef]
51. Driscoll, J.R.; Healy, D.M. Computing Fourier Transforms and Convolutions on the 2-Sphere. *Adv. Appl. Math.* **1994**, *15*, 202–250. [CrossRef]
52. Jaynes, E.T. Information Theory and Statistical Mechanics. *Phys. Rev.* **1957**, *106*, 620. [CrossRef]
53. Hastie, T.J.; Tibshirani, R.J. *Generalized Additive Models*; Chapman and Hall/CRC: Boca Raton, FL, USA, 1990.
54. Fahrmeir, L.; Kneib, T.; Lang, S.; Marx, B. *Regression: Models, Methods and Applications*; Springer: Berlin/Heidelberg, Germany, 2013.
55. Berger, A.L.; Della Pietra, V.J.; Della Pietra, S.A. A maximum entropy approach to natural language processing. *Comput. Linguist.* **1996**, *22*, 39–71.
56. Ivezić, Ž.; Connolly, A.J.; VanderPlas, J.T.; Gray, A. *Statistics, Data Mining, and Machine Learning in Astronomy*; Princeton University Press: Princeton, NJ, USA, 2014.
57. Ng, A.Y.; Jordan, M.I. On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. *Adv. Neural Inf. Process. Syst.* **2002**, *14*, 841–884.
58. Eilers, P.H.C.; Marx, B.D. Flexible smoothing with B-splines and penalties. *Stat. Sci.* **1996**, *11*, 89–121. [CrossRef]
59. Duchon, J. Splines minimizing rotation-invariant semi-norms in Sobolev spaces. In *Constructive Theory of Functions of Several Variables*; Springer: Berlin/Heidelberg, Germany, 1977; pp. 85–100. [CrossRef]
60. Wahba, G. Spline interpolation and smoothing on the sphere. *SIAM J. Sci. Stat. Comput.* **1981**, *2*, 5–16. [CrossRef]
61. Wood, S.N. Thin plate regression splines. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2003**, *65*, 95–114. [CrossRef]
62. Li, Y.; Ye, Z. Boosting in Univariate Nonparametric Maximum Likelihood Estimation. *IEEE Signal Process. Lett.* **2021**, *28*, 623–627. [CrossRef]
63. Yanagihara, H.; Ohtaki, M. On Avoidance of the Over-fitting in the B-Spline Non-parametric Regression Model. *Jpn. J. Appl. Stat.* **2004**, *33*, 51–69. [CrossRef]

64. Coolen, A.C.C.; Sheikh, M.; Mozeika, A.; Aguirre-Lopez, F.; Antenucci, F. Replica analysis of overfitting in generalized linear regression models. *J. Phys. A* **2020**, *53*, 365001. [CrossRef]
65. Tikhonov, A.N.; Arsenin, V.Y. *Solutions of Ill-Posed Problems*; V. H. Winston & Sons: Washington, DC, USA; New York, NY, USA, 1977; p. xiii+258.
66. Wood, S.N. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2011**, *73*, 3–36. [CrossRef]
67. Marra, G.; Wood, S.N. Practical variable selection for generalized additive models. *Comput. Stat. Data Anal.* **2011**, *55*, 2372–2387. [CrossRef]
68. Bailer-Jones, C.A.L. Bayesian inference of stellar parameters and interstellar extinction using parallaxes and multiband photometry. *Mon. Not. R. Astron. Soc.* **2011**, *411*, 435–452. [CrossRef]
69. Patterson, H.D.; Thompson, R. Recovery of inter-block information when block sizes are unequal. *Biometrika* **1971**, *58*, 545–554. [CrossRef]
70. Kelsall, J.E.; Diggle, P.J. Non-parametric estimation of spatial variation in relative risk. *Stat. Med.* **1995**, *14*, 2335–2342. [CrossRef]
71. Caticha, A. Entropic Inference. In *Proceedings of the Bayesian Inference and Maximum Entropy Methods in Science and Engineering*; American Institute of Physics Conference Series; Mohammad-Djafari, A., Bercher, J.F., Bessiére, P., Eds.; AIP Publishing: Melville, NY, USA, 2011; Volume 1305, pp. 20–29. [CrossRef]
72. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.