Statistical Inference for the Statistical Isotropy Stochastic Gravitational Wave Background

Hongru Zhao

School of Statistics, University of Minnesota

SGWB Research Collaborators

Interdisciplinary collaboration across Statistics and Physics

University of Minnesota

School of Statistics

- Sara Algeri
- Galin L. Jones
- Xiangyu Zhang
- Jesse Miller

School of Physics and Astronomy

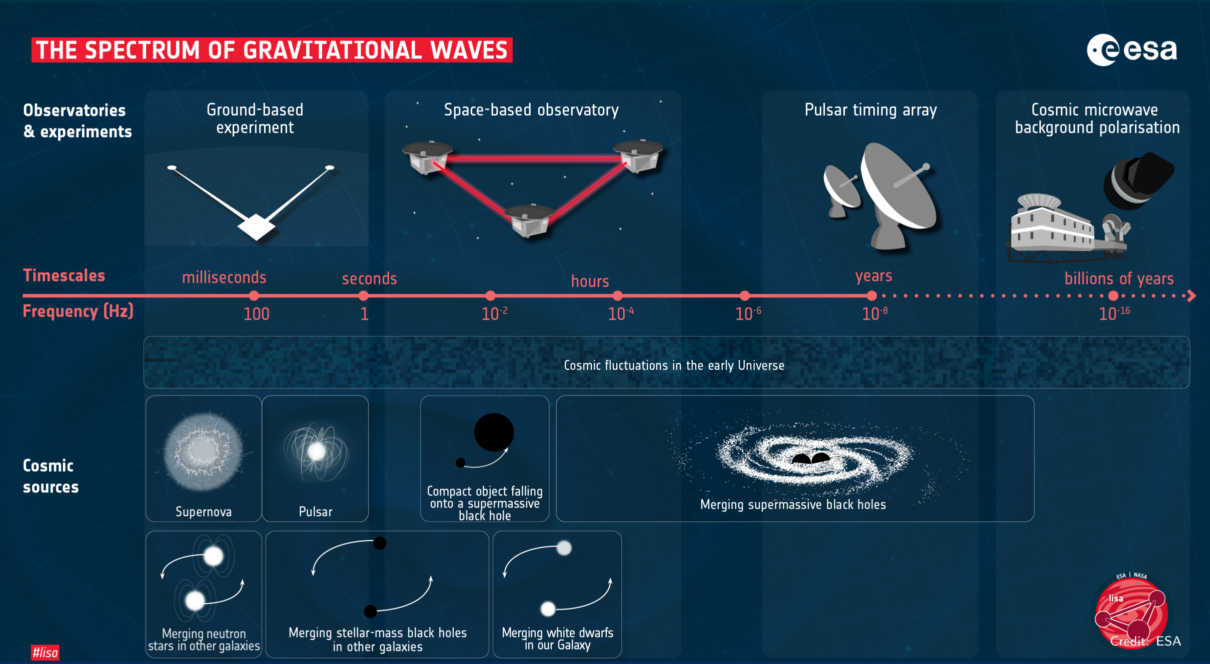
- Vuk Mandic
- Erik Floden

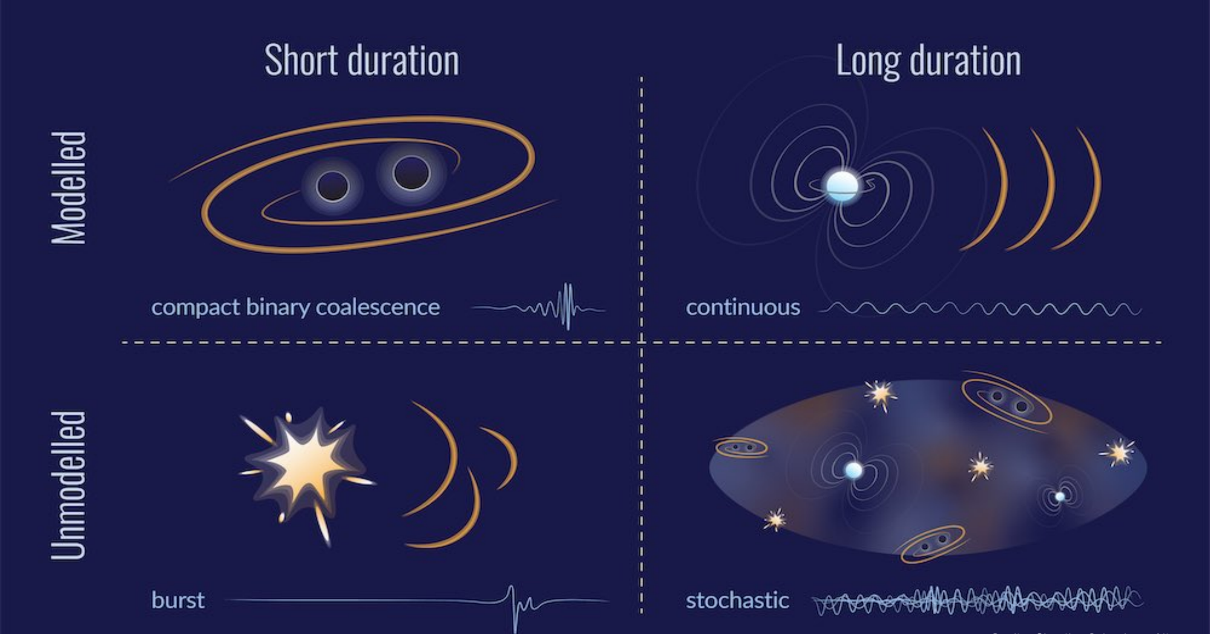
External Collaborators

University of Texas Rio Grande Valley
Department of Physics and Astronomy

- Joseph D. Romano
- Deepali Agarwal

Statistical methodology for gravitational-wave inference


THE SPECTRUM OF GRAVITATIONAL WAVES
esa
Observatories & experiments
Ground-based experiment
Space-based observatory
Pulsar timing array
Cosmic microwave background polarisation
Timescales
milliseconds
seconds
hours
years
billions of years
Frequency (Hz)
100
1
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-16}
Cosmic fluctuations in the early Universe
Cosmic sources
Supernova
Pulsar
Compact object falling onto a supermassive black hole
Merging supermassive black holes
Merging neutron stars in other galaxies
Merging stellar-mass black holes in other galaxies
Merging white dwarfs in our Galaxy
#lisa
ESA | NASA
lisa
Credit: ESA



Credit: Shanika Galaudage, Nice

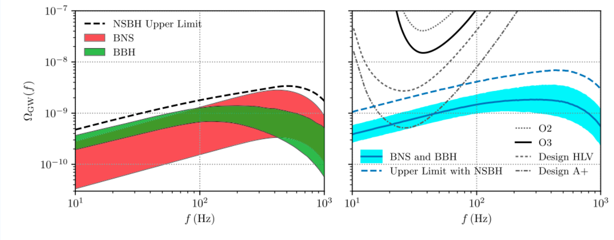
Stochastic Gravitational Wave Background (SGWB)

- **Origin:** Incoherent superposition of unresolvable/sub-threshold astrophysical GW sources + events in early universe.
- **Statistical assumptions:** Gaussian, stationary in time and space, unpolarized, anisotropic.
- **Quantification:**

$$\left\langle \tilde{h}_A(f, \hat{n})\, \tilde{h}^*_{A'}(f', \hat{n}') \right\rangle = \delta_{AA'}\delta(f - f')\delta^2(\hat{n}, \hat{n}')\, P(f, \hat{n})$$

Predictions for Detection of Monopole Component

$$\Omega_{\rm gw}(f) = \frac{f}{\rho_c}\frac{d\rho_{\rm gw}}{df} \propto f^3 P(f)$$

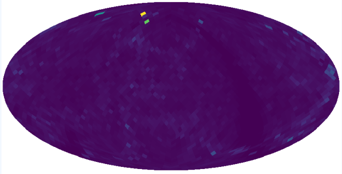


$$P(f, \hat{n}) = P(f)$$

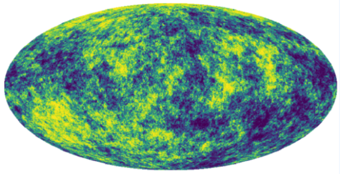
Phy. Rev. D 104, 022004 (2021)

Anisotropy: beyond one number

Isotropic
same power in every direction



Anisotropic
power varies across sky directions



Statistical Isotropy
fluctuation power summarized by the angular power spectrum

$$P(\hat{n}) = \sum_{\ell m} a_{\ell m} Y_{\ell m}(\hat{n}), \qquad A_\ell = \frac{1}{2\ell + 1} \sum_{m=-\ell}^{\ell} |a_{\ell m}|^2.$$

Here ℓ indexes angular scale, and A_ℓ is the angular power spectrum.

Cusin+ PRD 96, 103019 (2017); Chung+ PRD 106, 082005 (2022); arxiv 2202.05105

GW radiometer: mapping anisotropy

Cross-correlation principle

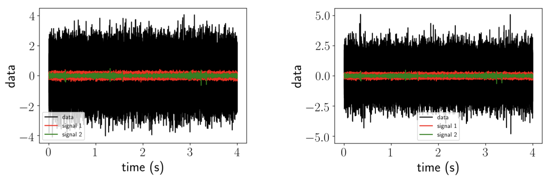
$$d_1 = h + n_1, \qquad d_2 = h + n_2.$$

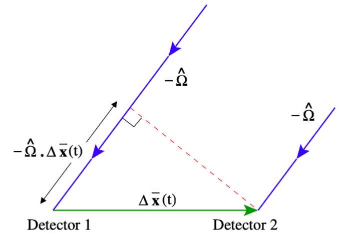
A true GW background induces correlated detector response, while detector noise is assumed uncorrelated.

Cross spectral density

$$C(t, f) = \tilde{d}_1^*(t, f)\tilde{d}_2(t, f).$$

Its expectation is a detector-weighted projection of the sky.





arxiv 1909.00269; Ballmer 2006; Mitra+ 2008; Thrane+ 2009

GW radiometer: mapping anisotropy

- GW signal is assumed to be correlated, but detector noise is not.

$$\left\langle \widehat{C}_{12} \right\rangle = \langle d_1 d_2 \rangle = \langle h^2 \rangle + \langle h n_1 \rangle + \langle h n_2 \rangle + \langle n_1 n_2 \rangle$$

Signal and noise uncorrelated

Noise uncorrelated

$$\left\langle \underbrace{C(t;f)}_{\text{Observation}} \right\rangle = \left\langle \tilde{d}_1^*(t;f)\tilde{d}_2(t;f) \right\rangle \propto \underbrace{\gamma_{\ell m}(t,f)}_{\text{Detector's response}} \; \underbrace{P_{\ell m}(f)}_{\text{Source}}$$

arxiv 1909.00269; Ballmer 2006; Mitra+ 2008; Thrane+ 2009

Detector geometry and overlap reduction function

The overlap reduction function describes how the detector pair geometry affects the sensitivity to a common stochastic gravitational wave signal.

- **Detector rotation:** different orientations of the detectors change how they respond to the GW signal.
- **Detector separation:** the distance between detectors introduces time delays and phase differences.
- **GW arrival direction:** waves from different directions project differently onto the detectors.

Roadmap: two views of the same coefficients

Fixed effect view

$a_{\ell m}$ = unknown realized sky coefficients

$$A_\ell = \frac{1}{2\ell + 1} \sum_{m=-\ell}^{\ell} |a_{\ell m}|^2.$$

Random effect view

$$a_{\ell m} \sim \mathcal{CN}(0, A_\ell)$$

$$p(C \mid A_\ell) = \int p(C \mid a) p(a \mid A_\ell)\, da$$

Target is to estimate free A_ℓ and infer θ in $A_\ell(\theta)$.

Viewpoint I: fixed effects

Separable frequency–sky model

$$P(f,\hat{n}) = H(f)P(\hat{n})$$

$$P(\hat{n}) = \sum_{\ell=0}^{\ell_{\max}} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\hat{n})$$

This definition introduces the spherical harmonic coefficients $a_{\ell m}$.

$a_{\ell m}$: fixed effect parameters

$$C(t;f) \mid a \sim \mathcal{CN}(\mu(t;f), N(t;f))$$

$$\mu(t;f) = \sum_{\ell=0}^{\ell_{\max}} \sum_{m=-\ell}^{\ell} H(f)\gamma_{\ell m}(t;f)a_{\ell m}$$

$$\mathrm{Cov}\big(C(t;f), C(t';f') \mid a\big) \approx \delta_{tt'}\delta_{ff'}P_I(t;f)P_J(t;f)$$

$P_I(t;f)$ and $P_J(t;f)$ are the one-sided power spectral densities of the two detector outputs.

Fixed effect likelihood and the clean map MLE

Compact notation Let $\alpha = (\ell, m)$, and define

$$B_{tf,\alpha} = H(f)\gamma_\alpha(t;f), N_{tf} = P_I(t;f)P_J(t;f).$$

Then

$$\mu_{tf} = \sum_\alpha B_{tf,\alpha} a_\alpha .$$

Gaussian log-likelihood

$$-2\ell(a) = \sum_{t,f} \frac{\left|C(t;f) - \sum_\alpha B_{tf,\alpha} a_\alpha\right|^2}{N_{tf}} + \text{const.}$$

$$X_\alpha = \sum_{t,f} B^*_{tf,\alpha} N^{-1}_{tf} C(t;f)$$

$$\Gamma_{\alpha\beta} = \sum_{t,f} B^*_{tf,\alpha} N^{-1}_{tf} B_{tf,\beta}$$

$$\widehat{a}_\alpha = (\Gamma^{-1} X)_\alpha$$

X is the **dirty map** and Γ is the **Fisher matrix**. Because the model is linear Gaussian, the MLE has the closed form solution $\widehat{\boldsymbol{a}} = \Gamma^{-1} X$.

Credit: Thrane et al., Phys. Rev. D **80**, 122002 (2009).

From $\widehat{a}_{\ell m}$ to a moment unbiased $\widehat{A}_\ell$

Distributional result: From the Gaussian likelihood,

$$X \mid a \sim \mathcal{CN}(\Gamma a, \Gamma),$$

so

$$\widehat{a} \mid a \sim \mathcal{CN}(a, \Gamma^{-1}).$$

Angular power spectrum

$$A_\ell := \frac{1}{2\ell+1} \sum_m |a_{\ell m}|^2.$$

$$\mathbb{E}\left(|\widehat{a}_{\ell m}|^2 \mid a\right) = |a_{\ell m}|^2 + (\Gamma^{-1})_{\ell m, \ell m}$$

$$\widehat{A}_\ell = \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} \left(|\widehat{a}_{\ell m}|^2 - (\Gamma^{-1})_{\ell m, \ell m}\right)$$

$$\mathbb{E}(\widehat{A}_\ell \mid a) = A_\ell.$$

Exact conditional covariance of $\widehat{A}_\ell$

$$\mathrm{Cov}\left(\widehat{A}_\ell, \widehat{A}_{\ell'} \mid a\right) = \sum_{m,m'} \frac{1}{(2\ell+1)(2\ell'+1)} \left[\left|(\Gamma^{-1})_{\ell m,\ell' m'}\right|^2 + 2\,\mathrm{Re}\left\{a^*_{\ell m}(\Gamma^{-1})_{\ell m,\ell' m'}\, a_{\ell' m'}\right\}\right]$$

Plug-in covariance estimate $\widehat{\Sigma}_{\ell\ell'} = \mathrm{Cov}\left(\widehat{A}_\ell, \widehat{A}_{\ell'} \mid \widehat{a}\right)$. In practice, replace the unknown $a_{\ell m}$ in the formula by its MLE $\widehat{a}_{\ell m}$.

This covariance matrix is used to construct the weight matrix for fitting parametric angular power spectrum models by generalized least squares.

Credit: Phys. Rev. D **113**, 023051 (2026).

From free A_ℓ to a parametric model $A_\ell(\theta)$

Step 1: free angular power estimate

1. Estimate the clean map coefficients $a_{\ell m}$ by $\widehat{a}_{\ell m}$.
2. Construct free realized angular powers $\widehat{A}_\ell$.
3. Use the plug-in covariance estimate $\widehat{\Sigma}$.

Step 2: parametric angular power fit

If one imposes structure, replace the free angular powers by $A_\ell = A_\ell(\theta)$.
The covariance estimate $\widehat{\Sigma}$ gives the generalized least squares weight.

$$\widehat{\theta} = \arg\min_{\theta}\{\widehat{A} - A(\theta)\}^T \widehat{\Sigma}^{-1}\{\widehat{A} - A(\theta)\}.$$

This is a two step generalized nonlinear least squares estimator: first estimate $a_{\ell m}$ and $\widehat{A}_\ell$; then use $\widehat{\Sigma}$ to fit θ. No assumption $a_{\ell m} \sim \mathcal{CN}(0, A_\ell)$ is used.

Fixed effect model validation without a full distribution

Null hypothesis

$$H_0 : \quad \mathbb{E}(\widehat{A}) = A(\theta_0)$$

We test the mean angular power model after estimating θ_0.

Why not just use Gaussian likelihood?

$\widehat{A}_\ell$ is built from squared complex Gaussian clean map coefficients. Its exact distribution is non Gaussian and depends on the covariance structure.

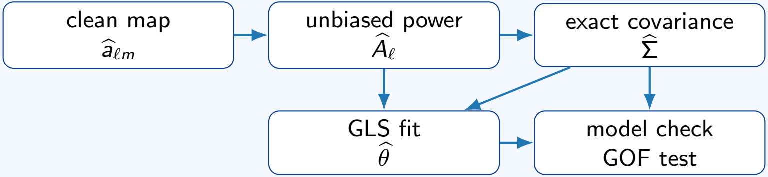
$$\widehat{\epsilon} = \widehat{\Sigma}^{-1/2}\{\widehat{A} - A(\widehat{\theta})\}$$

$$w_{k,N} = \frac{1}{\sqrt{N}} \sum_{i=1}^{k} \widehat{\epsilon}_i$$

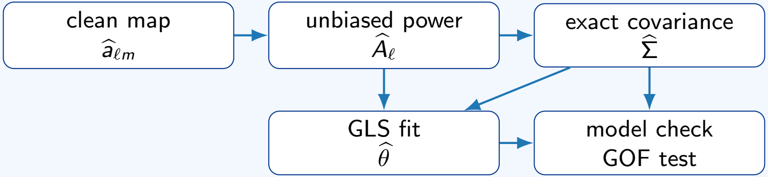
$$\widetilde{K}_N = \max_k |v_{k,N}|, \qquad \widetilde{C}_N = \frac{1}{N} \sum_k |v_{k,N}|^2$$

The transformation removes the fitted model drift, giving a distribution free goodness of fit test under broad conditions.

Fixed effect workflow summary



Fixed effect workflow summary



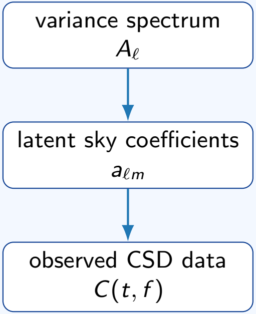
$$\boxed{\text{No assumption } a_{\ell m} \sim \mathcal{CN}(0, A_\ell) \text{ is used here.}}$$

Viewpoint II: random effects

$$a_{\ell m} \sim \mathcal{CN}(0, A_\ell)$$

$$A_\ell = \mathrm{Var}(a_{\ell m})$$

Here A_ℓ is not the squared amplitude of one realized sky. It is a population variance parameter for the random coefficients.



Random effects: the sky coefficients are unobserved latent variables.

From complex CSDs to a real mixed model

Conditional complex CSD model

$$C(t;f) \mid a \sim \mathcal{CN}(\mu(t;f), N(t;f))$$

$$\mu(t;f) = \sum_{\ell=0}^{\ell_{\max}} \sum_{m=-\ell}^{\ell} H(f)\gamma_{\ell m}(t;f)a_{\ell m}$$

$$\mathrm{Cov}\big(C(t;f), C(t';f') \mid a\big) \approx \delta_{tt'}\delta_{ff'}P_I(t;f)P_J(t;f)$$

Why realify? Standard Gaussian likelihood and mixed model algorithms are easiest to write for real vectors.

$$y = x\beta + Z\xi + \varepsilon$$

$$\varepsilon \sim N(0,R), \qquad \xi \sim N(0,G_A).$$

β is the isotropic/monopole amplitude. The vector ξ contains the independent real degrees of freedom from $a_{\ell m}$.

Random effect covariance for spherical harmonics

Real degrees of freedom For $m = 0$:

$$\xi_{\ell 0} = a_{\ell 0} \sim N(0, A_\ell)$$

For $m > 0$:

$$\xi_{\ell m}^{\mathrm{Re}} = \mathrm{Re}(a_{\ell m}), \xi_{\ell m}^{\mathrm{Im}} = \mathrm{Im}(a_{\ell m}) \sim N\left(0, \frac{A_\ell}{2}\right)$$

$$G_A = \mathrm{blockdiag}\left(A_1 D_1, A_2 D_2, \ldots, A_{\ell_{\max}} D_{\ell_{\max}}\right)$$

$$D_\ell = \mathrm{diag}\left(1, \frac{1}{2}, \ldots, \frac{1}{2}\right)$$

Each A_ℓ controls the variance of all modes at angular scale ℓ.

Marginalization: remove the unobserved sky coefficients

Joint density

$$p(y, \xi \mid \beta, A) = p(y \mid \xi, \beta) p(\xi \mid A)$$

Observed data likelihood

$$p(y \mid \beta, A) = \int p(y \mid \xi, \beta) p(\xi \mid A) \, d\xi$$

$$y \mid \beta, A \sim N(x\beta, V_A)$$

$$V_A = R + Z G_A Z^T$$

Basic statistical point: marginalization averages over all possible latent skies according to the model $a_{\ell m} \sim \mathcal{CN}(0, A_\ell)$.

One-step MLE for angular power parameters

Unrestricted variance spectrum

$$(\widehat{\beta}, \widehat{A}_1, \ldots, \widehat{A}_{\ell_{\max}}) = \arg\max_{\beta \geq 0,\, A_\ell \geq 0} \ell(\beta, A; y)$$

Parametric variance spectrum

$$A_\ell = A_\ell(\theta), (\widehat{\beta}, \widehat{\theta}) = \arg\max_{\beta \geq 0, \theta} \ell\{\beta, A(\theta); y\}$$

$$A_\ell(a, \lambda) = a\ell^\lambda$$

Contrast with fixed effects Fixed effects: estimate $\widehat{A}_\ell$ first, then fit θ.
Random effects: estimate θ directly from the marginal likelihood.

A sufficient statistic: the augmented dirty map

Fisher–Neyman factorization theorem

A statistic $T(y)$ is sufficient for θ if the likelihood can be written as

$$p(y \mid \theta) = h(y)\, g_\theta\{T(y)\},$$

where $h(y)$ is free of θ.

$$\mathcal{X}_\mathbb{R} = \begin{pmatrix} s_{xy} \\ u_y \end{pmatrix} = \begin{pmatrix} x^T W y \\ Z^T W y \end{pmatrix}, \quad W = R^{-1}$$

$$\mathcal{F}_\mathbb{R} = \begin{pmatrix} s_{xx} & u_x^T \\ u_x & M \end{pmatrix} = \begin{pmatrix} x^T W x & x^T W Z \\ Z^T W x & Z^T W Z \end{pmatrix}.$$

Exact factorization for the marginal model

For $y \mid \theta_A \sim N(x\beta, R + Z G_A Z^T)$,

$$p(y \mid \theta_A) = h(y)\, g_{\theta_A}(\mathcal{X}_\mathbb{R}; \mathcal{F}_\mathbb{R}).$$

Let

$$H_A = I + G_A^{1/2} M G_A^{1/2}, \quad \mathcal{A}_A = G_A^{1/2} H_A^{-1} G_A^{1/2}.$$

$$g_{\theta_A} = |H_A|^{-1/2} \exp\Big[\tfrac{1}{2} u_y^T \mathcal{A}_A u_y + \beta\{s_{xy} - u_x^T \mathcal{A}_A u_y\} \\ - \tfrac{1}{2}\beta^2\{s_{xx} - u_x^T \mathcal{A}_A u_x\}\Big].$$

Therefore, for fixed $\mathcal{F}_\mathbb{R}$, $\mathcal{X}_\mathbb{R}$ is sufficient for $\theta_A = (\beta, A_1, \ldots, A_{\ell_{\max}})$. Here u_y is the real dirty map and M is the real Fisher block.

Likelihood model hierarchy for testing

M1: noise only $\beta = 0,\ A_\ell = 0$

↓ add monopole

M2: isotropic SGWB $\beta \geq 0,\ A_\ell = 0$

↓ add random anisotropy

M3: power-law random anisotropy $\beta \geq 0,\ A_\ell = a\ell^\lambda$

↓ relax spectrum shape

M4: unrestricted angular variance $\beta \geq 0,\ A_\ell \geq 0$

Nested marginal likelihoods give likelihood-ratio tests. Because variance components lie on the boundary $A_\ell \geq 0$, we calibrate the LR statistic by parametric bootstrap rather than relying on a standard χ^2 reference law.

Likelihood-ratio test: definition

Nested hypotheses

$$H_0 : \theta \in \Theta_0, \qquad H_1 : \theta \in \Theta_1 - \Theta_0, \qquad \Theta_0 \subset \Theta_1.$$

Let

$$L(\theta) = p(y \mid \theta), \qquad \ell(\theta) = \log L(\theta).$$

The null model is the more restricted model; the alternative model has extra signal parameters.

MLEs under the two models

$$\widehat{\theta}_0 = \arg\max_{\theta \in \Theta_0} \ell(\theta), \qquad \widehat{\theta}_1 = \arg\max_{\theta \in \Theta_1} \ell(\theta).$$

$$\Lambda_{\mathrm{LR}} = 2\{\ell(\widehat{\theta}_1) - \ell(\widehat{\theta}_0)\}$$

Decision rule Large Λ_{LR} means that the larger model improves the marginal likelihood after integrating out the latent coefficients $a_{\ell m}$.

$$\text{Reject } H_0 \quad \text{if} \quad \Lambda_{\mathrm{LR}} > c_\alpha.$$

In this SGWB problem, c_α is obtained by bootstrap because the variance parameters satisfy $A_\ell \geq 0$.

O3 data set in the marginal MLE framework

Power-law random-effect model

$$A_\ell = a\ell^{\lambda}, \qquad \lambda \in [-5, 0].$$

The same sufficient-statistic likelihood is applied to the O3 data product.

$$\widehat{\beta} = 0,\ \widehat{a} = 5.3680 \times 10^{-20},\ \widehat{\lambda} = 0.$$

Likelihood ratio detection test M1 vs M3

Statistic	Observed	p_{boot}	95% critical
LR	0.004277	0.775045	4.425050

Interpretation: the O3 fit is boundary dominated. The fitted monopole is zero, the fitted variance is extremely small, and the bootstrap calibrated LR test does not reject the noise only model.

References and figure credits

- Credit: ESA
- Credit: https://www.ligo-india.in/outreach/gravitational-waves/
- Credit: Shanika Galaudage; arXiv:1909.00269
- Ballmer 2006; Mitra+ 2008
- Thrane et al., Phys. Rev. D 80, 122002 (2009)
- Phys. Rev. D 104, 022004 (2021)
- Cusin+ PRD 96, 103019 (2017)
- Chung+ PRD 106, 082005 (2022); arXiv:2202.05105
- Phys. Rev. D 113, 023051 (2026)
- Burke-Spolaor, S. et al. Astron. Astrophys. Rev. 27, 5 (2019)
- Credit: Jishnu Suresh
- Romano and Cornish, Living Reviews in Relativity 20, 2 (2017)
- Allen and Romano, Phys. Rev. D 59, 102001 (1999)
- User manuscript: likelihood inference for angular-power variance components