

A novel approach to optimize clustering of parametric map-based component separation for upcoming CMB polarization satellites.

Wassim Kabalan ,* Josquin Errard ,[†] and Alexandre Boucaud
APC

+ SciPol team
APC

(SciPol Collaboration)

(Dated: May 6, 2025; Received; Revised; Accepted; Published)

We present a novel JAX-powered parametric component separation approach that optimizes the degrees of freedom in the component separation parametrization, achieving minimum variance after foreground cleaning. This work is the first to incorporate an algorithm for optimizing the clustering of sky regions where the frequency scaling laws are assumed constant. Our technique is particularly applied to the challenging large angular scales encountered in future CMB space missions—illustrated here with a LiteBIRD-like instrument targeting the measurement of primordial gravitational waves down to $r < 0.001$. Furthermore, our efficient and modular framework paves the way for the extension of traditional component separation techniques.

I. INTRODUCTION

Detecting the B-mode polarization in the Cosmic Microwave Background (CMB) is a central objective in modern cosmology, as it offers direct observational evidence of primordial gravitational waves generated during the universe’s inflationary period [1, 2]. The amplitude and shape of this B-mode signal are direct probes of inflationary energy scales, potentially reaching physics at $\sim 10^{16}$ GeV. Upcoming experiments, such as LiteBIRD [3], aim to measure these faint signals with unprecedented sensitivity, targeting constraints on the tensor-to-scalar ratio r down to levels of $r < 0.001$.

Achieving these ambitious goals is complicated by the presence of astrophysical foregrounds, primarily arising from Galactic synchrotron emission and thermal dust [4, 5]. These foregrounds dominate the CMB polarization signal across much of the sky and exhibit spatially varying spectral properties. Large angular scales (low multipoles) are both most sensitive to primordial B-modes and most affected by foregrounds and instrument systematics, making accurate separation especially critical in this regime.

Component separation methods can be broadly classified as parametric—assuming a model for the spectral energy distributions (SEDs) of foregrounds—or non-parametric, which rely on statistical independence or internal templates [6, 7]. Key tools include COMMANDER [8, 9], SMICA [7, 10], and FGBUSTER [11, 12], each providing different approaches to the trade-off between flexibility and tractability.

Our method builds on the parametric class, enhancing its flexibility by introducing spatial variability in a systematic, data-driven manner. Although many traditional parametric approaches assume spatial uniformity in spectral parameters, recent observations indicate that

the spectral energy distributions of foregrounds vary significantly across the sky [4, 5]. Frameworks such as FGBUSTER [11, 12], as used in [13], address this by supporting *clustered* parameter configurations—where the sky is divided into a finite number of regions (or *clusters*), each sharing a common set of spectral parameters. However, the computational cost of optimizing multiple such clustering choices in these frameworks is often prohibitive, limiting their practical flexibility.

In contrast, our framework is explicitly designed to efficiently explore a large space of clustering configurations. Our implementation, based on **FURAX** [14] (Chania et al. In Prep), can be viewed as a JAX-native generalization of FGBUSTER, built for end-to-end differentiability, GPU acceleration, and large-scale model selection. Leveraging the high-performance computing capabilities of JAX [15], FURAX enables scalable and reproducible component separation pipelines suited for next-generation satellite missions.

In the remainder of this paper, we detail our methodology (Section II), illustrate our framework’s advantages over existing techniques (Section V), and discuss implications for future observational missions (Section VI).

II. METHODOLOGY

A. Parametric Component Separation

We model multi-frequency sky observations using the standard parametric framework. The observed Stokes vector in each pixel is modeled as a linear combination of astrophysical components with additive (Gaussian) noise:

$$\mathbf{d} = \mathbf{A}(\boldsymbol{\beta}) \mathbf{s} + \mathbf{n}, \quad (1)$$

where:

- $\mathbf{d} \in \mathbb{R}^{N_d}$: observed data vector (frequency \times polarization \times pixels),
- $\mathbf{s} \in \mathbb{R}^{N_s}$: sky components at a reference frequency (polarization \times pixels),
- $\mathbf{A}(\beta) \in \mathbb{R}^{N_d \times N_s}$: mixing matrix encoding spectral dependencies,
- $\mathbf{n} \sim \mathcal{N}(0, \mathbf{N}) \in \mathbb{R}^{N_d}$: Gaussian noise with known covariance \mathbf{N} (frequency \times polarization \times pixels),

Each column of the mixing matrix $\mathbf{A}(\beta)$ models the spectral energy density of each component. In practice, this includes a modified blackbody (MBB) emission law for thermal dust and a power-law dependence for synchrotron radiation, with β denoting the set of corresponding spectral parameters ($\beta_{dust}, T_{dust}, \beta_{synchrotron}$).

Assuming Gaussian noise, the negative log-likelihood under this model is:

$$-2 \ln \mathcal{L}(\mathbf{s}, \beta) = (\mathbf{d} - \mathbf{As})^\top \mathbf{N}^{-1} (\mathbf{d} - \mathbf{As}) + \text{const.}, \quad (2)$$

To estimate the sky components \mathbf{s} , we solve:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{s}} \Big|_{\mathbf{s}=\hat{\mathbf{s}}} = 0, \quad (3)$$

which yields the generalized least squares solution:

$$\hat{\mathbf{s}} = (\mathbf{A}^\top \mathbf{N}^{-1} \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{N}^{-1} \mathbf{d}, \quad (4)$$

Substituting this into the likelihood (2) eliminates dependence on \mathbf{s} , giving the spectral likelihood [16]:

$$\ln \mathcal{L}_{\text{spec}}(\beta) \propto (\mathbf{A}^\top \mathbf{N}^{-1} \mathbf{d})^\top (\mathbf{A}^\top \mathbf{N}^{-1} \mathbf{A})^{-1} (\mathbf{A}^\top \mathbf{N}^{-1} \mathbf{d}), \quad (5)$$

which depends only on the spectral parameters β , and is used to optimize their values.

B. Cluster-Based Generalization

To account for spatial variability in foreground spectral parameters, we generalize the model (1) by allowing β to vary across the sky. However, assigning one parameter set per pixel would increase the statistical uncertainty of the recovered components—particularly the CMB and derived cosmological parameters [17]. To mitigate this, we introduce a clustered model in which the sky is divided into regions (or *clusters*) that share common spectral parameters.

We use *spherical K-means* clustering to partition the sky into disjoint regions $\{\mathcal{C}_k\}$, where:

$$\beta(\hat{n}) = \beta_k \quad \text{for all } \hat{n} \in \mathcal{C}_k,$$

thus assigning a single set of spectral parameters to each region. The clustering configuration is parameterized solely by two quantities: the random seed used for centroid initialization, and the total number of desired clusters.

Our implementation, available via `jax-healpy` [18], is adapted specifically for spherical coordinates and inspired by the `kmeans_radec` package [19]. Clustering operates on right ascension and declination coordinates, minimizing angular distances across the celestial sphere. Although input points are in spherical coordinates, centroid updates during K-means iterations are performed via 3D Cartesian averaging (over (x, y, z) coordinates mapped from (RA, DEC)) to ensure numerical stability and efficient convergence.

The entire clustering algorithm is implemented to be fully compatible with JAX’s functional programming paradigm and accelerator-based execution. It is fully vectorized, supports automatic differentiation (autodiff), and avoids explicit Python loops, making it highly scalable for large sky datasets and efficient within end-to-end optimization frameworks.

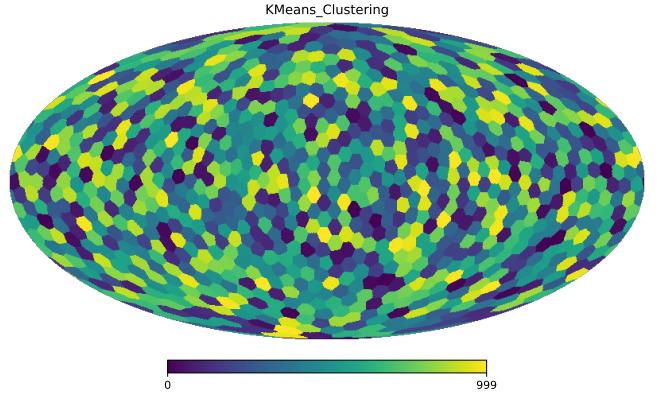


FIG. 1: Example of K-means clustering applied to a Planck dust map, showing regions with approximately uniform spectral behavior. Each color indicates a distinct cluster.

In practice, we initialize the clustering of sky pixels using predefined templates such as the Planck Galactic masks (e.g., GAL020, GAL040) [20]. These masks divide the sky into regions with different levels of Galactic foreground contamination—typically distinguishing low-, medium-, and high-foreground areas based on Galactic latitude or emission intensity thresholds. Each clustering configuration—defined by the number of patches used for parameters like β_d , T_d , and β_s —specifies a distinct model within the data framework of Equation (1). Rather than optimizing the clustering assignments directly, we evaluate a grid of such configurations by computing the spectral likelihood in Equation (5) for each, selecting the one that minimizes the variance of the recovered CMB.

C. Grid Search over Clustering Models

Our model allows the clustering of different spectral parameters (e.g., dust and synchrotron) on independent sets of sky patches, without requiring them to align spa-

tially. The number of clusters assigned to each spectral parameter is treated as a discrete modeling choice, assumed specifically for the purpose of component separation. For example, one configuration might use 100 patches for β_d , 20 patches for T_d , and 10 patches for β_s .

Letting \mathcal{G} denote the space of possible clustering configurations:

$$\mathcal{G} = \{K_{\beta_d}\} \times \{K_{T_d}\} \times \{K_{\beta_s}\},$$

we perform a structured grid search across \mathcal{G} , evaluating each configuration $\{\mathcal{C}_k\} \in \mathcal{G}$ by fitting spectral parameters β and reconstructing the CMB component. Specifically, we maximize the spectral likelihood:

$$\mathcal{L}_{\text{spec}}(\beta, \{\mathcal{C}_k\}),$$

where the likelihood depends on both the parameter values and the spatial clustering structure.

Maximizing the spectral likelihood ensures a good fit to the observed frequency maps. However, it does not by itself prevent overfitting, especially when increasing the number of free parameters by introducing finer spatial patching. In low signal-to-noise regions, overly flexible models can fit noise realizations rather than true sky signals, increasing statistical residuals in the reconstructed CMB map.

To mitigate this, we define a secondary selection metric—based on the variance of the reconstructed CMB across noise realizations (described in Section II E). The CMB variance acts as a proxy for statistical residuals: minimizing it leads to reduced noise-induced contamination and more robust recovery of the underlying cosmological signal. Thus, the optimal clustering configuration is chosen to balance spectral fit quality with minimization of statistical uncertainty.

Each configuration is assessed over multiple noise realizations and sky regions, ensuring that the selected model is both statistically stable and robust against noise-driven biases.

Given the large size of the configuration space and the need to evaluate multiple noise realizations and sky regions, we developed a distributed, parallel evaluation framework, described next.

D. Distributed and Parallel Execution

To evaluate large grids of clustering configurations across multiple noise realizations and sky regions, we developed a distributed optimization engine: `jax-grid-search` [21]. This JAX-native framework enables high-throughput exploration of parametric component separation models at scale, fully leveraging modern GPU architectures.

As defined previously, the grid \mathcal{G} spans combinations of patch counts across spectral parameters.

Each element in this grid represents a distinct clustering configuration. For every such configuration, we eval-

uate the component separation pipeline across multiple noise realizations and sky regions.

To handle this computational load efficiently, we combine two forms of parallelism:

- **Intra-device parallelism** via `jax.vmap`, enabling batched execution of component separation fits on a single GPU.
- **Inter-device parallelism** using MPI-style slicing: the global parameter grid is evenly partitioned across all processes, regardless of whether the number of combinations is divisible by the number of workers.

Formally, for P total processes, each process $I \in [0, P - 1]$ is assigned a contiguous chunk of the global grid:

$$\mathcal{G}_I = \mathcal{G} \left[\left\lfloor \frac{I \cdot N}{P} \right\rfloor : \left\lfloor \frac{(I + 1) \cdot N}{P} \right\rfloor \right], \quad N = |\mathcal{G}|$$

This ensures robust workload distribution, load balancing, and fault tolerance. It avoids assumptions about grid divisibility and gracefully handles incomplete or interrupted runs—making it well suited for large-scale HPC deployments.

Each batch is independently evaluated, with results checkpointed to disk for aggregation and analysis. This modular design supports recovery, progress tracking, and reproducibility—all essential for exploratory model selection pipelines.

E. Clustering Objective: Variance Minimization

For each clustering configuration in our grid, we perform maximum likelihood estimation of the spectral parameters:

$$\forall \mathcal{C}_k \in \{\mathcal{C}\} : \quad \beta_k^* = \arg \max_{\beta_k} \mathcal{L}_{\text{spec}}(\beta_k, \mathcal{C}_k) \quad (6)$$

Using these parameters, we reconstruct the sky components via:

$$\hat{\mathbf{s}} = (\mathbf{A}(\beta_k^*)^\top \mathbf{N}^{-1} \mathbf{A}(\beta_k^*))^{-1} \mathbf{A}(\beta_k^*)^\top \mathbf{N}^{-1} \mathbf{d} \quad (7)$$

$$= W(\mathbf{d}) \quad (8)$$

From the reconstructed components, we extract the CMB signal \hat{s}_{CMB} . We then evaluate its variance across noise realizations:

$$\sigma_{\text{CMB}}^2 = \langle \text{Var}[\hat{s}_{\text{CMB}}(\hat{n})] \rangle_{\text{noise}} \quad (9)$$

Although the spectral parameters β are defined in a clustered fashion—varying across spatial regions—the reconstructed components $\hat{\mathbf{s}}$, and hence \hat{s}_{CMB} , are evaluated as a full-sky solution.

This variance serves as the loss function for selecting the best clustering configuration. It acts as a proxy for residual foreground contamination and directly impacts the uncertainty on the inferred tensor-to-scalar ratio r .

Minimizing this variance reduces both statistical uncertainty and systematic leakage in the B -mode power spectrum. As shown in [22], residuals from imperfect foreground modeling propagate into biases on C_ℓ^{BB} , making this metric critical for achieving robust and unbiased cosmological constraints.

F. The FURAX Framework

PAPER IN PREP

To support this method, we introduce FURAX [14]—a modular JAX-powered framework to express and optimize parametric component separation likelihoods.

FURAX is built around a central goal: to express symbolic algebraic expressions (e.g., likelihoods) directly in code using modular, composable operators. These operators abstract common structures like the mixing matrix, noise weighting, or parameter dispatching, and can be combined into scalable computation graphs.

Key features:

- Differentiable algebraic operators for the *mixing matrix* $\mathbf{A}(\beta)$, the *noise diagonal* operator \mathbf{N}^{-1} , and cluster-wise *SED parameter dispatch* using mask-based routing.
- Conjugate gradient solvers via the Lineax library [23] to solve inverse systems efficiently without explicitly forming matrices.
- Fully vectorized execution across simulation ensembles and clustering grids, leveraging JAX [15] for efficient large-scale computation.

All core computations in FURAX are implemented using memory-efficient *linear operators*, avoiding explicit dense matrices. This symbolic and modular approach allows FURAX to express and optimize large, realistic models without the memory bottlenecks associated with explicit matrix construction.

In the scope of this paper, we assume isotropic white noise and matched beam resolution. Although advanced treatments (e.g., beam convolution, correlated noise) exist in the literature, they are not our focus here. However, FURAX is designed to make such extensions straightforward in future developments, including applications to map-making, half-wave plate systematics, atmospheric effects, and beyond.

The framework integrates natively with the JAX ecosystem and supports optimization and inference workflows through libraries such as Optax [24] and NumPyro [25].

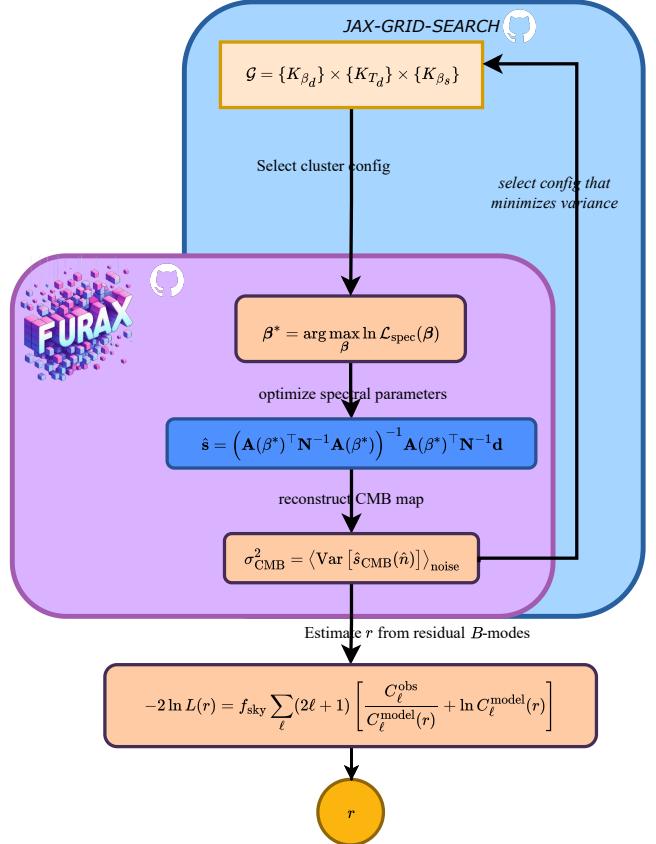


FIG. 2: Overview of the FURAX pipeline for component separation and tensor-to-scalar ratio estimation. The process begins by selecting a clustering configuration from the discrete grid $\mathcal{G} = \{K_{\beta_d}\} \times \{K_{T_d}\} \times \{K_{\beta_s}\}$, evaluated in parallel using JAX-GRID-SEARCH. Each configuration undergoes spectral parameter optimization and map reconstruction using FURAX. The variance of the recovered CMB map is used as a selection metric to identify the best configuration. The final CMB map is then used to compute the B -mode power spectrum and infer the tensor-to-scalar ratio r .

III. SPECTRAL LIKELIHOOD MAXIMIZATION

The central task in parametric component separation is the estimation of spectral parameters β by maximizing the *spectral likelihood* $\mathcal{L}_{\text{spec}}(\beta)$ (5). This likelihood serves as the objective function that drives parameter inference across the model.

Thanks to JAX’s automatic differentiation, gradients of this objective with respect to all free parameters are computed efficiently. This applies even in spatially varying models, where spectral parameters are defined per cluster and must be dispatched to the appropriate sky pixels. Crucially, this indexing and dispatching operation, performed via binary masks, is itself implemented

in a differentiable manner. As a result, the likelihood function is compatible with gradient-based optimization.

FURAX performs spectral likelihood maximization using gradient-based optimization from the Optax library [24]. The default method is **L-BFGS**, a limited-memory quasi-Newton optimizer that leverages curvature information from past gradients. Ideally, optimization would proceed via a full Newton step:

$$\theta_{n+1} = \theta_n - H^{-1} \nabla L(\theta_n), \quad (10)$$

where H is the Hessian of the loss. However, computing and inverting this matrix becomes intractable for large-scale clustered models with thousands of parameters. L-BFGS avoids this cost by constructing a low-rank approximation using a limited history of gradients and parameter updates, enabling efficient and memory-scalable updates.

In practice, we observe that L-BFGS in FURAX exhibits superior **stability and convergence** (see Figure 3) compared to earlier approaches such as FG-BUSTER. This is primarily due to two factors: (i) we use exact gradients from automatic differentiation in JAX, ensuring consistent and smooth optimization dynamics; and (ii) we avoid the use of Singular Value Decomposition (SVD) [26], which FG-BUSTER employs as a performance shortcut at the cost of numerical noise in gradients. These design choices make FURAX robust to initialization and capable of converging reliably even in high-dimensional, highly clustered configurations.

We also experimented with **TNC** (Truncated Newton with Constraints) [27], as used in FG-BUSTER. In our framework, TNC is accessed through `jaxopt` [28], which wraps `scipy.optimize.minimize` [29]. However, this implementation runs exclusively on CPU and is incompatible with JAX’s `jit` compilation or `vmap` vectorization. More critically, **TNC does not scale well with parameter count**—as the number of clusters increases, the optimization problem becomes stiff and memory-bound, leading to significantly longer runtimes and poorer convergence behavior.

First-order optimizers such as **Adam** [30] were also tested, but proved to be ill-suited for this application. Adam applies a uniform learning rate schedule across all parameters, yet in component separation, the physical scales of spectral parameters differ substantially: for instance, β_d varies between 0.5 and 5, T_d between 10 and 50 K, and β_s between –6 and –1.

Moreover, the spectral likelihood typically reaches values on the order of 10^8 (for example, at `nside` 64), making fixed or scheduled learning rates ineffective. In contrast, L-BFGS employs a line search strategy that dynamically adapts the step size to local curvature, enabling stable convergence even in high-dimensional, multi-scale optimization settings.

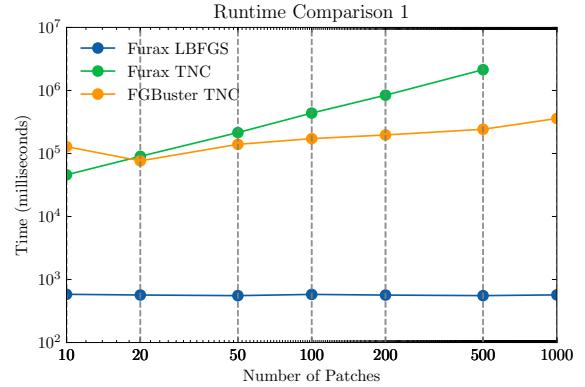


FIG. 3: Runtime comparison of FG-BUSTER (TNC), FURAX (TNC), and FURAX (L-BFGS) as a function of the number of clusters. L-BFGS scales robustly with parameter count and consistently outperforms TNC in clustered models.

IV. TENSOR-TO-SCALAR RATIO ESTIMATION

We estimate the tensor-to-scalar ratio r from the B -mode power spectrum of the reconstructed CMB maps. The process involves modeling residual contamination, combining it with signal templates, and performing likelihood-based inference of r [31, 32].

a. *BB Spectrum Convention.* All angular power spectra are reported in the standard units:

$$\text{coeff} = \frac{\ell(\ell+1)}{2\pi},$$

b. *Systematic and Statistical Residuals.* To quantify contamination in the recovered CMB, we separate residuals into two components:

- **Systematic residual power** is obtained by re-running the full component separation pipeline on foreground-only skies d_{ds} , which exclude both CMB and instrumental noise. This evaluates the imprint of foreground leakage that cannot be filtered out by the model:

$$C_\ell^{\text{syst}} = C_\ell(W(d_{\text{ds}})),$$

where W denotes the reconstruction operator as defined in (7). The resulting residuals capture systematic foreground contamination intrinsic to the modeling assumptions and patch structure.

- **Total residual power** is computed from the full reconstruction error across noise realizations:

$$C_\ell^{\text{res}} = \left\langle C_\ell \left(\hat{s}_{\text{CMB}}^{(n)} - s_{\text{CMB}}^{\text{true}} \right) \right\rangle_n,$$

where n indexes the noise realizations and $\hat{s}_{\text{CMB}}^{(n)} \equiv W(d^{(n)})$. This represents the total error power,

including residual foregrounds and instrumental noise.

- **Statistical residual power**, due solely to instrumental noise, is extracted by subtraction:

$$C_\ell^{\text{stat}} = C_\ell^{\text{res}} - C_\ell^{\text{syst}}.$$

This decomposition isolates stochastic uncertainty from systematic bias, and C_ℓ^{stat} is used as the effective noise floor for likelihood estimation.

c. BB Spectrum Model. We model the theoretical B -mode spectrum as:

$$C_\ell^{BB}(r) = r \cdot C_\ell^{\text{tensor}} + C_\ell^{\text{lensing}} + C_\ell^{\text{stat}},$$

where C_ℓ^{tensor} is a primordial tensor-mode template normalized to $r = 1$, C_ℓ^{lensing} is the fixed lensing B -mode contribution, and C_ℓ^{stat} accounts for statistical residuals from noise and imperfect foreground cleaning.

d. Likelihood and Inference. The tensor-to-scalar ratio r is inferred by maximizing a chi-squared likelihood function under the assumption of Gaussian-distributed residuals:

$$-2 \ln \mathcal{L}_{\text{cosmo}}(r) = f_{\text{sky}} \sum_\ell (2\ell+1) \left[\frac{C_\ell^{\text{obs}}}{C_\ell^{\text{model}}(r)} + \ln C_\ell^{\text{model}}(r) \right]$$

We evaluate this function over a grid of r values and report the best-fit \hat{r} corresponding to the minimum of the negative log-likelihood. Uncertainties are estimated from the curvature of the likelihood around its maximum.

V. RESULTS

A. Validation of the Framework

To assess the performance of our clustered component separation framework, we construct a synthetic validation setup in which the true sky components and spectral parameters are fully known. This controlled environment enables precise measurement of reconstruction quality, residual contamination, and estimator consistency.

The simulation begins by defining a known sky signal \mathbf{s} in Equation (1), consisting of Stokes Q/U maps for CMB, thermal dust, and synchrotron emission, all specified at a common reference frequency. This sky vector is then mixed using a known mixing matrix $\mathbf{A}(\beta)$, where β encodes the ground-truth spectral parameters for each component. Table I summarizes the simulation parameters.

The noise vector \mathbf{n} is sampled as Gaussian white noise, scaled to match LiteBIRD-like polarization depth across frequency channels. Each clustering configuration is evaluated across 100 independent noise realizations to reduce statistical variance and assess estimator stability.

We perform a grid search over discrete cluster configurations (defined in Section II C), varying the number of

patches assigned to each spectral parameter. For each setup, we run the full component separation pipeline by minimizing the spectral negative log-likelihood. The resulting CMB map \hat{s}_{CMB} is then compared with the input to assess reconstruction fidelity and the accuracy of tensor-to-scalar ratio estimation.

TABLE I: Validation setup used in the synthetic benchmark.

Parameter	Value / Range
Resolution	$N_{\text{side}} = 64$
Sky mask	GAL020
Noise realizations	100
Patch count grid	$\beta_d \in [10, 300]$, $T_d, \beta_s \in [5, 20]$
True Patch count	$\beta_d = 100$, $T_d = 20K$, $\beta_s = 20$

Figure 4 shows the grid search results. The configuration minimizing the CMB variance coincides with the configuration yielding the lowest spectral negative log-likelihood, indicating that improved physical reconstruction aligns with optimal statistical fit. Importantly, the recovered clustering matches the true simulation setup exactly, validating the robustness of the variance-based selection criterion against noise-driven overfitting.

B. Comparison Setup: Grid-Search K-means Clustering vs. Multi-Resolution

We compare two spatial modeling strategies for parametric component separation, both implemented within the FURAX framework:

- **K-means clustering**, which uses spherical K-means to learn spatial patches over which spectral parameters are shared (introduced in Section II B).
- **Multi-resolution grouping**, which defines parameter patches by downgrading HEALPix maps to coarser resolutions, following the strategy of [13].

Although applied to the same sky and instrument setup, these approaches differ fundamentally in how they capture spatial variability. The K-means strategy allows a flexible, data-driven exploration of spatial granularity via grid search, while the multi-resolution method imposes predefined patch sizes. This comparison evaluates reconstruction fidelity, residual suppression, and consistency in inferred cosmological parameters.

Simulation Setup

Simulated skies are generated using PySM3 [33–35], combining CMB, thermal dust, and synchrotron emission (`c1d1s1` model). Observations are modeled after a LiteBIRD-like instrument, including polarization-sensitive Gaussian noise across multiple frequency bands.

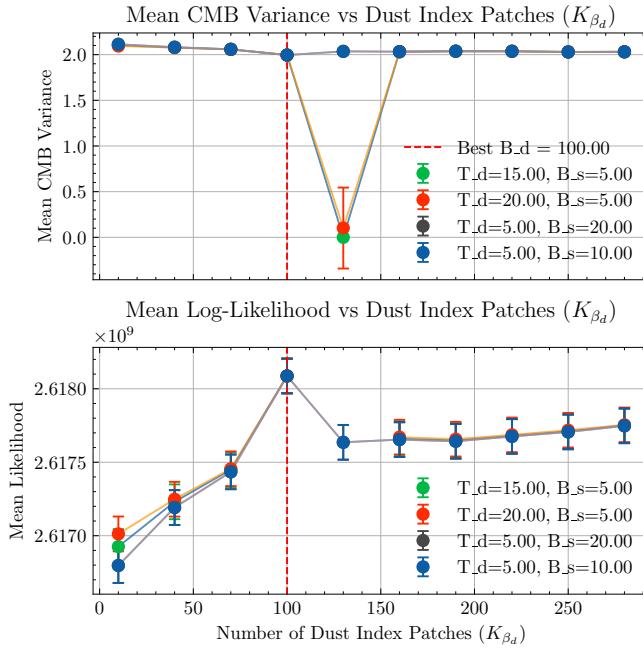


FIG. 4: Grid search results for varying the number of dust spectral index patches K_{β_d} , while keeping other parameters fixed. Each curve corresponds to a different configuration of K_{T_d} and K_{β_s} . **Top:** Mean variance of the reconstructed CMB component across noise realizations, used as the primary selection metric. **Bottom:** Mean spectral negative log-likelihood, offset for readability. The vertical red line marks the true value used in the simulated sky ($K_{\beta_d} = 100$), and the dashed lines indicate the global extrema of each metric. The configuration with $K_{\beta_d} = 100$, $K_{T_d} = 15$, and $K_{\beta_s} = 5$ was used to generate the reference clustered map. **Minimizing the CMB variance successfully recovers the true clustering configuration, demonstrating the effectiveness of variance-based model selection in the presence of noise.**

TABLE II: Simulation parameters used for both methods.

Parameter	Value
Sky model	c1d1s1
Instrument	LiteBIRD-like sensitivity
Map resolution	$N_{\text{side}} = 64$
Noise model	White Gaussian noise
Noise realizations	100 per region

Sky Region Partitioning

To probe performance across different foreground regimes, we define distinct sky zones using Planck-based Galactic masks:

- **K-means clustering** is applied to six disjoint sub-

regions, obtained by splitting each mask into upper and lower hemispheric parts: GAL020_U, GAL020_L, GAL040_U, GAL040_L, GAL060_U, and GAL060_L.

- **Multi-resolution grouping** is applied separately to the full-sky masks: GAL020, GAL040, and GAL060.

These masks span from high Galactic latitudes (low foregrounds, GAL020) to regions closer to the Galactic plane (higher foregrounds, GAL060).

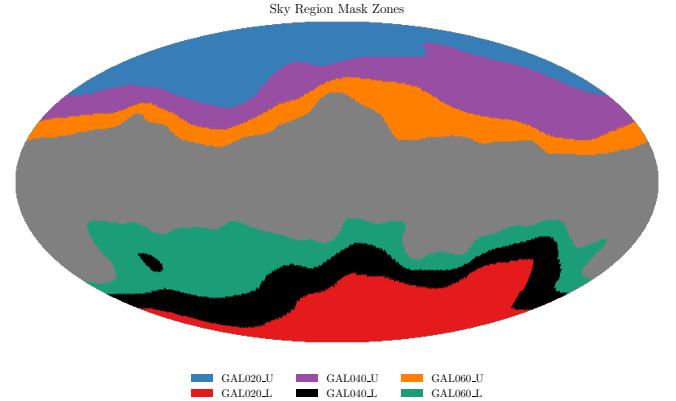


FIG. 5: Sky regions used in the comparison study. Colored areas correspond to the six disjoint regions for adaptive clustering; black outlines indicate full-sky masks used for multi-resolution grouping.

c

Grid-Search Clustering

In the adaptive clustering approach, spectral parameters are assigned to pixel clusters defined via spherical K-means. A structured grid search over patch counts is performed, selecting the configuration that minimizes the CMB variance across 100 noise realizations. This fully data-driven search avoids assumptions about spatial scales and allows optimal patch sizes to emerge from the data itself.

TABLE III: Grid-search clustering configuration.

Parameter	Patch Count Range
β_d	100 to 5000 (in steps of 100)
T_d, β_s	{1, 5, 20, 30, 50, 60, 70, 80}
Sky regions	Six zones (GAL020_U, etc.)
Noise realizations	100 per region

Multi-Resolution Grouping

The multi-resolution grouping method creates spatial patches by degrading the original HEALPix pixelization

to coarser N_{side} values, separately for each spectral parameter. After patching, the maps are upsampled back to $N_{\text{side}} = 64$ to match the data grid.

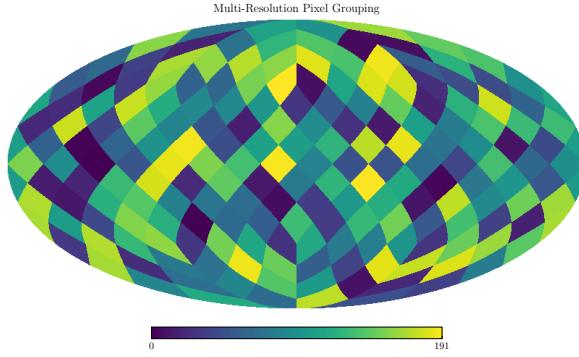


FIG. 6: Multi-resolution mega-pixel structure. Example showing downgraded patch regions from $N_{\text{side}} = 64$ to $N_{\text{side}} = 8$.

The patching resolutions vary depending on the sky mask and spectral parameter, as summarized in Table IV.

TABLE IV: Multi-resolution configuration per mask region.

Region (Mask)	$ \beta_d $	$ T_d $	$ \beta_s $
GAL060 (low lat.)	64	8	4
GAL040 (mid lat.)	64	4	2
GAL020 (high lat.)	64	0	2

While computationally efficient, this fixed-resolution approach lacks the flexibility to adapt to localized foreground complexity, which explains the superior reconstruction performance achieved by adaptive clustering.

C. Runtime and Parallelization

The extensive clustering and component separation scans were performed on the *Jean Zay* supercomputer, utilizing NVIDIA A100 GPUs distributed across 4 compute nodes (32 GPUs total). Computations were executed using the `jax-grid-search` framework (Section II D), leveraging JAX’s native parallelization capabilities for high-throughput evaluation.

The full campaign encompassed:

- 3200 clustering configurations,
- each evaluated across 100 independent noise realizations,
- over 6 distinct sky regions (Figure 5),

leading to approximately 1.92 million independent component separation runs.

Despite the large computational volume, the full scan completed in under 30 hours, consuming approximately 1000 GPU-hours. For comparison, earlier CPU-based frameworks such as FGBUSTER required around 40 minutes per configuration, making such exhaustive parameter scans practically infeasible.

This performance gain stems from two key design choices: (i) extensive batching and vectorization of likelihood evaluations within each GPU, and (ii) distributed slicing of the grid across multiple devices without assuming divisibility. The implementation supports checkpointing, failure recovery, and fully reproducible execution, ensuring robustness for large HPC deployments.

Thanks to this scalability, FURAX enables a new regime of data-driven model selection, where clustering hypotheses are tested systematically across noise realizations and sky regions, rather than being fixed a priori.

D. CMB Reconstruction Comparison

We assess the CMB reconstruction performance achieved by the two spatial modeling strategies introduced in Section VB: clustering-based patching via K-means and fixed-resolution multi-resolution grouping. Both methods are applied to the same simulated sky (CMB plus foregrounds) with LiteBIRD-like noise and angular resolution.

The analysis focuses on the accuracy of the reconstructed CMB polarization fields, examining both Stokes Q and U components, and on quantifying residual foreground leakage.

Patch Structures

Before presenting reconstruction results, we first illustrate the spatial patch layouts produced by each method. Figures 7 and 8 show the pixel clustering for the three spectral parameters: β_d , T_d , and β_s .

The K-means clustering approach (Figure 7) produces flexible, irregular patches that adapt to the local sky complexity by minimizing the CMB variance. In contrast, the multi-resolution method (Figure 8) relies on a fixed HEALPix downgrading scheme, imposing uniform patch sizes across large sky regions.

These structural differences in patching strongly influence the quality of component separation, as discussed in the following sections.

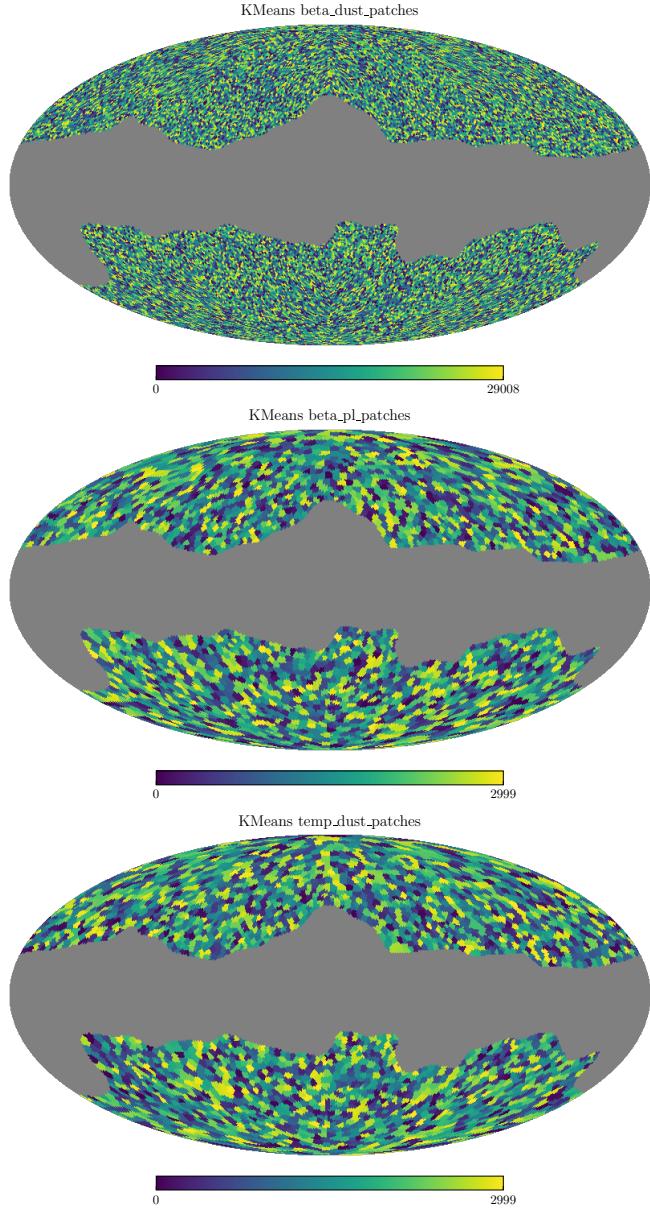


FIG. 7: Final FURAX clustering configuration. Top to bottom: patches for β_d , T_d , and β_s , obtained by grid search minimizing the CMB reconstruction variance. Clusters are irregular and data-driven, adapting to spatial variations in foreground complexity.

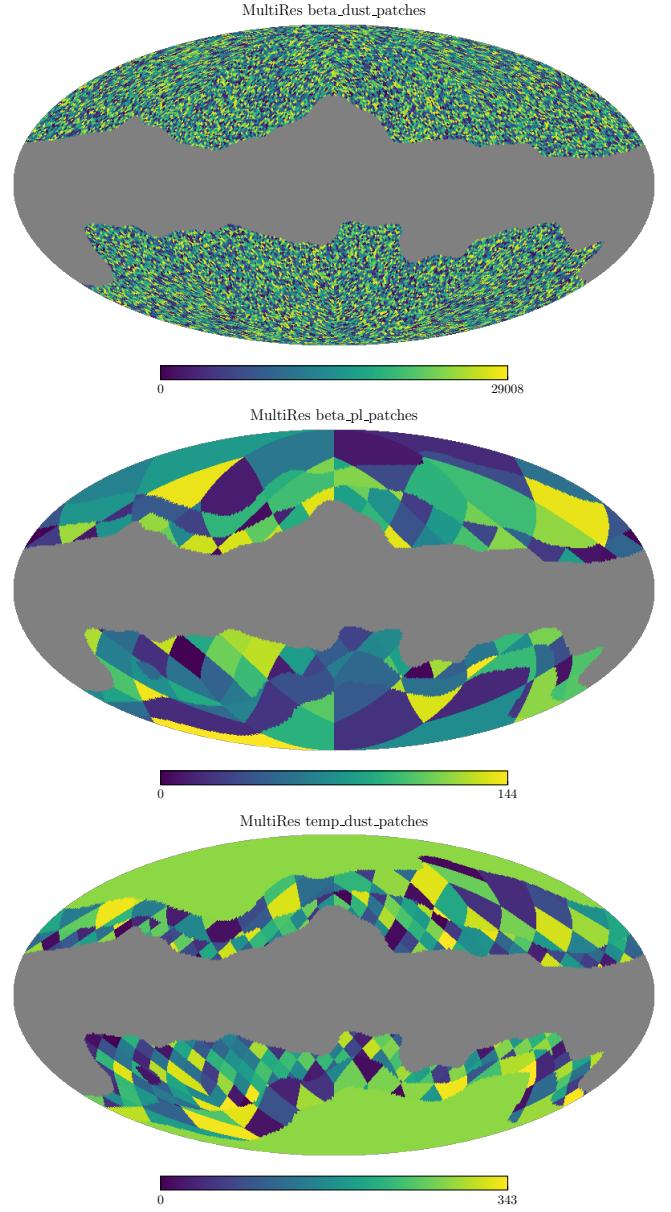


FIG. 8: Multi-resolution patch structure. Top to bottom: groupings for β_d , T_d , and β_s , generated by downgrading HEALPix maps to fixed N_{side} resolutions. Patches are regular and uniform, independent of the true sky variability.

E. Residuals and CMB Reconstruction

We assess the quality of CMB reconstruction achieved by the three spatial modeling strategies: single-patch, multi-resolution, and K-means clustering. The evaluation focuses separately on:

- Residual CMB polarization maps (Q and U)
- Residual B -mode power spectra

Residual CMB Polarization Maps

We first examine the residual maps, defined as the difference between the reconstructed and true CMB polarization fields for Stokes Q and U .

Figure 9 compares the results across methods.

The multi-resolution approach achieves the visually cleanest maps, with minimal foreground residuals across the sky. K-means clustering yields slightly larger residuals but still substantially improves over the single-patch baseline, which shows significant large-scale errors. These results illustrate the benefits of spatially adaptive modeling; however, as we will discuss in the next section, lower visual residuals do not necessarily imply better statistical performance at the power spectrum level.

Residual B-Mode Power Spectra

We now decompose the residuals into systematic and statistical contributions at the level of the B -mode angular power spectrum.

Figure 10 shows the residual spectra across all multipoles. To contextualize the scale of residual contamination, we also plot the theoretical B -mode power spectra from lensing and primordial tensor modes ($r = 1$), computed using the CAMB python package [36].

The single-patch model suffers from large systematic residuals (C_ℓ^{syst}), exceeding the total residuals C_ℓ^{res} at many scales. This counter-intuitive behavior results from the averaging procedure defined in IV: noise fluctuations partially cancel when averaging over noise realizations, while coherent foreground leakage persists, leading to an excess of systematic power.

The multi-resolution method suppresses systematic residuals but at the cost of increased statistical noise C_ℓ^{stat} , reflecting its greater modeling freedom and risk of noise overfitting.

In contrast, K-means clustering achieves the best overall performance, balancing systematic suppression and statistical stability to minimize the total residual power across the observed multipole range.

This confirms that minimizing the CMB variance through adaptive clustering results in a globally improved reconstruction, even if the residual maps appear slightly noisier compared to multi-resolution grouping.

F. Tensor-to-Scalar Ratio Estimation

We now assess the impact of residual foreground contamination on the estimation of the tensor-to-scalar ratio r . For each spatial modeling strategy—K-means clustering, multi-resolution grouping, and single global patch—we compute the r -likelihood using the reconstructed B -mode power spectra.

Figure 12 shows the resulting normalized likelihood distributions aggregated over sky regions and noise realizations.

The K-means clustering approach yields the most accurate and precise measurement, with a likelihood distribution narrowly centered around the true r value. The multi-resolution method exhibits a slight positive bias and broader dispersion, consistent with increased statistical noise in the reconstruction. In contrast, the single-patch model produces a deceptively tight constraint but a significant bias, strongly overestimating r due to residual systematic contamination. This behavior reflects the dominance of coherent residuals, which shift the likelihood peak while artificially reducing the inferred uncertainty. These trends mirror those observed in the residual B -mode spectra and further demonstrate the importance of flexible spatial modeling to achieve unbiased and robust cosmological inference.

a. Summary of Results.

- **K-means clustering** achieves the tightest and most unbiased constraints on r .
- **Multi-resolution grouping** shows slight bias and wider uncertainty due to statistical noise amplification.
- **Single global patch** yields a biased estimate of r despite a narrow likelihood, driven by uncorrected systematic residuals.

G. Impact of Noise on Clustered Component Separation

An important advantage of our approach lies in its adaptability to noise realization variability. In a multi-resolution approach, the spatial patch structure—defined by a fixed HEALPix downgrading—is predetermined and remains constant across noise realizations. This rigidity prevents the model from adapting to noise-induced fluctuations in the data.

In contrast, our K-means clustering strategy combined with a grid search allows the patch configuration to be dynamically re-optimized for each noise realization. For any given realization, the clustering grid search evaluates multiple spatial partitionings and selects the one that minimizes the variance of the recovered CMB map. This flexibility enables the model to better mitigate noise-driven distortions in the spectral parameter estimates, particularly in low signal-to-noise regions.

As a result, the adaptive clustering approach can reduce residual foreground contamination and enhance the fidelity of component separation by effectively tailoring the patch structure to each specific noise scenario. This noise-aware reconfiguration leads to lower residual power and improves robustness in downstream cosmological inference.

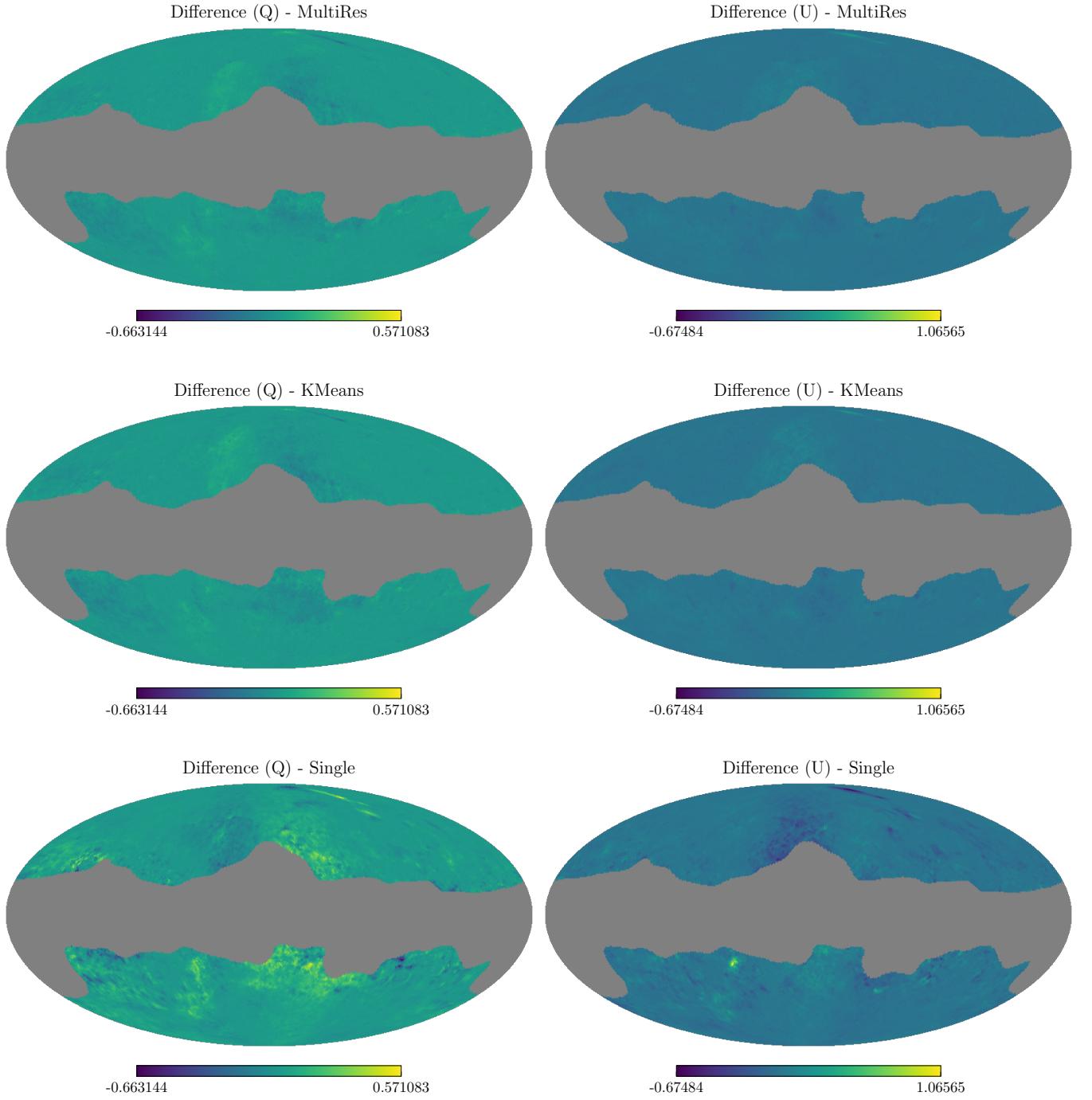


FIG. 9: Residual CMB polarization maps (Stokes Q and U) for single-patch, multi-resolution, and K-means clustering methods. While multi-resolution yields the lowest mean squared error (MSE)—with $MSE_Q = 6.7 \times 10^{-4}$, $MSE_U = 4.6 \times 10^{-4}$ —K-means performs only slightly worse ($MSE_Q = 8.0 \times 10^{-4}$, $MSE_U = 5.5 \times 10^{-4}$) but substantially outperforms the single-patch model. The small sacrifice in pixel-level reconstruction accuracy is offset by improved control of residual foregrounds—consequently giving a less biased estimation of the tensor-to-scalar ratio r .

VI. DISCUSSION

This work introduces a high-resolution, grid-based clustering approach for spatially varying parametric component separation, implemented in the modular and scalable FURAX framework. Our objective was to evaluate whether learned spectral patch structures can offer advantages over fixed-resolution strategies, using an exhaustive scan over clustering configurations compared against a baseline multi-resolution method.

Foreground Modeling at Scale

Rather than fixing patch scales by hand, we perform an *exhaustive grid search* over the number of patches per spectral parameter. This enables *data-driven selection* of spatial granularity, letting each sky region adopt the level of modeling complexity most appropriate to its foreground structure, as quantified by the variance in the recovered CMB map.

The resulting clusters, generated using *spherical K-means*, are highly flexible: patches are irregular, non-contiguous, and vary in size, adapting to the statistical structure of the input data. This avoids arbitrary binning and allows for more precise foreground modeling, particularly in regions with sharp spectral gradients or complex emission.

In contrast, the *multi-resolution approach* uses HEALPix downgrading to define fixed NSIDE patch maps. While fast and easy to implement, it lacks the capacity to adapt to regional sky complexity and yields higher residuals in zones with rapidly varying spectra. This difference highlights the value of learned patch structures that reflect the physical reality of the sky.

Computational Scalability and Pipeline Design

To support this large-scale scan—**over 2 million component separation evaluations** across 6 sky zones and 100 noise realizations per region—we leveraged *distributed GPU parallelism* using `jax-grid-search`. The entire sweep was completed in under 30 hours on 32 A100 GPUs across 4 nodes on the Jean Zay supercomputer.

This was made possible by the scalable design of FURAX, which includes:

- **JAX-native backend** with XLA compilation for performance
- **Checkpointing and recovery support** for distributed jobs
- **Composable operator design** for future extensions

Previous pipelines, including those used in [13], could not scale to this volume: a single multi-patch likeli-

hood fit could take up to 40 minutes, making exhaustive searches prohibitive. FURAX brings this class of analysis within reach, enabling principled exploration of spatial model structures at scale.

Reproducibility and Open Tools

All results in this paper were produced using *open-source software* developed or co-developed by the author, including the `furax` framework and `jax-grid-search`. The full codebase used to generate the simulations, run the experiments, and produce the plots is publicly available at:

<https://github.com/ASKabalan/furax-compsep-paper>

This ensures **transparent and reproducible research**, allowing others to easily re-run, extend, or adapt the pipeline for future use cases.

Limitations and Future Work

While effective, our current strategy is based on a *brute-force grid scan*. This could be made more efficient using modern *hyperparameter optimization* techniques such as Tree-structured Parzen Estimators or Bayesian search.

Other directions for future work include:

- Scaling to higher resolutions (e.g., $N_{\text{side}} = 128$ or 256)
- Adding priors on patch structure (e.g., smoothness or contiguity)
- Exploring disjoint and irregular clustering with parameter sharing
- Integrating additional systematic effects such as beam asymmetries, gain drift, or correlated noise

These extensions are naturally supported by the modular FURAX architecture.

Advanced Search Strategies for Cluster Configuration

The choice of cluster configuration in our framework is better understood as part of a *model selection problem*, rather than a standard parameter estimation task. It determines the structure of the likelihood model itself—how spectral parameters are spatially grouped and constrained.

We experimented with more advanced search techniques to optimize cluster configuration, including the `Optuna` [37] framework and discrete sampling strategies

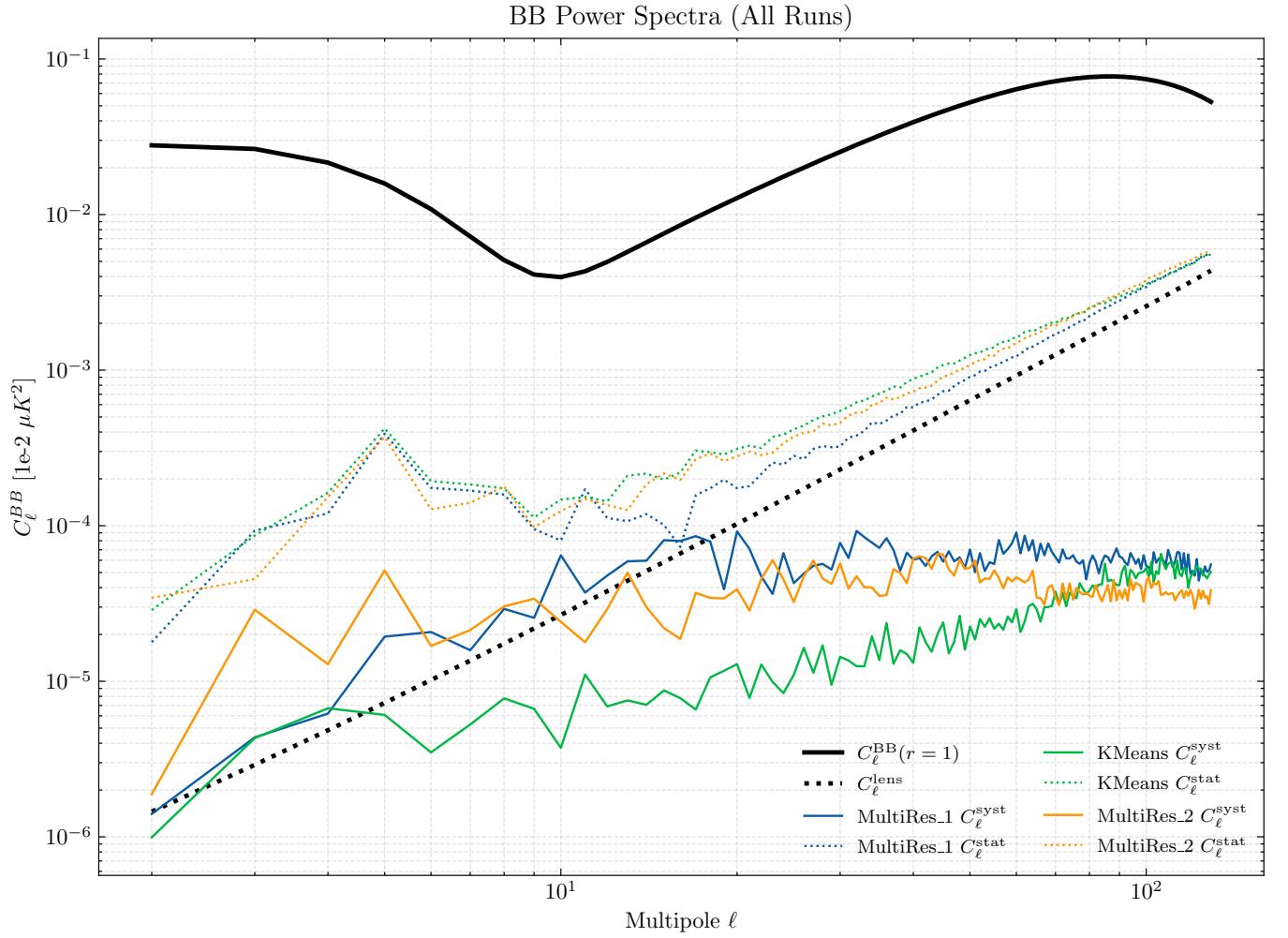


FIG. 10: Residual B -mode power spectra for single-patch, multi-resolution, and K-means clustering methods. The MultiRes-1 case (*cf.* Table IV, Fig. 8) illustrates the limited adaptability of pixel downgrading: moving from $N_{\text{side}} = 16$ to $N_{\text{side}} = 32$ entails a large, discrete increase in patch count, preventing intermediate configurations. This rigidity leads to overshooting, especially when increasing the patch granularity for T_d and β_s (MultiRes-2). In contrast, K-means clustering (Fig. 7) enables fine-grained, data-driven tuning of spatial complexity, yielding lower residuals by balancing systematic leakage and statistical variance.

like Gibbs sampling. However, these methods showed limited improvements over exhaustive scanning, likely due to the discrete and highly non-smooth nature of the objective landscape.

Future efforts could explore:

- Structured search methods that exploit correlations across sky zones
- Relaxed parameterizations that allow for differentiable optimization

For this work, the grid-based distributed search provided a robust and parallelizable solution, allowing exhaustive exploration of the model space in a tractable amount of time.

VII. CONCLUSIONS

This work presents a scalable, flexible framework for spatially varying parametric component separation, demonstrating the advantages of data-driven spectral patch selection through exhaustive clustering scans.

Compared to fixed-resolution baselines which were previously fine-tuned by hand, looking directly at the (in principle inaccessible) residuals, our approach yields:

- Lower residual contamination in the reconstructed CMB maps
- Reduced bias and tighter constraints on the tensor-to-scalar ratio r

- Greater adaptability to complex Galactic foreground structures

By optimizing spectral patching using a physical criterion—the variance of the recovered CMB—we show that adaptive clustering improves both reconstruction fidelity and cosmological parameter estimation.

The FURAX framework, fully open-source and scalable to millions of likelihood evaluations, provides a reproducible foundation for future high-fidelity CMB experiments targeting primordial B -mode polarization.

ACKNOWLEDGMENTS

The author would like to thank Josquin Errard for his guidance and supervision throughout this project. Special thanks to Wuhyun Sohn for his extensive and thoughtful reviews, and to Benjamin Beringue for his valuable assistance with mathematical formulations.

This work was supported by the SCIPOL project (scipol.in2p3.fr), funded by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (Principal Investigator: Josquin Errard, Grant agreement No. 101044073).

The computations were performed using HPC resources provided by the *Jean Zay* supercomputer at IDRIS under allocation 2024-AD010414161R2 granted by GENCI.

* Contact author: wassim@apc.in2p3.fr

† Contact author: josquin@apc.in2p3.fr

- [1] M. Kamionkowski, A. Kosowsky, and A. Stebbins, Detecting gravitational waves from inflation through cmb polarization, *Phys. Rev. Lett.* **78**, 2058 (1997).
- [2] U. Seljak and M. Zaldarriaga, Measuring polarization in the cosmic microwave background, *Phys. Rev. Lett.* **78**, 2054 (1997).
- [3] M. e. a. Hazumi, Litebird: a satellite for the studies of b-mode polarization and inflation from cosmic background radiation detection, *Progress of Theoretical and Experimental Physics* **2022**, 04A102 (2022).
- [4] Planck Collaboration, Planck 2015 results. x. diffuse component separation: Foreground maps, *Astron. Astrophys.* **594**, A10 (2016).
- [5] A. Meisner and D. Finkbeiner, Modeling thermal dust emission with planck and iras, *Astrophys. J.* **798**, 88 (2015).
- [6] J. e. a. Delabrouille, A full sky, low foreground, high resolution cmb map from wmap, *MNRAS* **346**, 1089 (2003).
- [7] J.-F. e. a. Cardoso, Component separation with flexible models. application to cmb data, *IEEE J. Sel. Top. Signal Process.* **2**, 735 (2008).
- [8] H. K. e. a. Eriksen, Joint bayesian component separation and cmb power spectrum estimation, *ApJ* **676**, 10 (2008).
- [9] Planck Collaboration, Planck 2015 results. x. diffuse component separation: Foreground maps, *A&A* **594**, A10 (2016).
- [10] Planck Collaboration, Planck 2015 results. ix. diffuse component separation: Cmb maps, *Astron. Astrophys.* **594**, A9 (2016).
- [11] R. Stompor and J. Errard, Bayesian component separation and cmb b-mode detection in the presence of instrumental systematics, *Phys. Rev. D* **97**, 083006 (2018).
- [12] J. Errard, Fgbuster: A flexible parametric component separation tool (2021), <https://github.com/fgbuster/fgbuster>.
- [13] e. a. Ally, Probing cosmic inflation with the litebirdcosmic microwave background polarization survey, *Progress of Theoretical and Experimental Physics* **2023**, 10.1093/ptep/ptac150 (2022).
- [14] P. Chanial and the CMBSciPol team, Furax: Flexible unified regression analysis for component separation, <https://github.com/CMBSciPol/furax> (2024), accessed: 2024-04-26.
- [15] J. Bradbury, R. Frostig, P. Hawkins, M. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. Vander-Plas, S. Wanderman-Milne, and Q. Zhang, Jax: composable transformations of python+numpy programs (2018), <http://github.com/google/jax>.

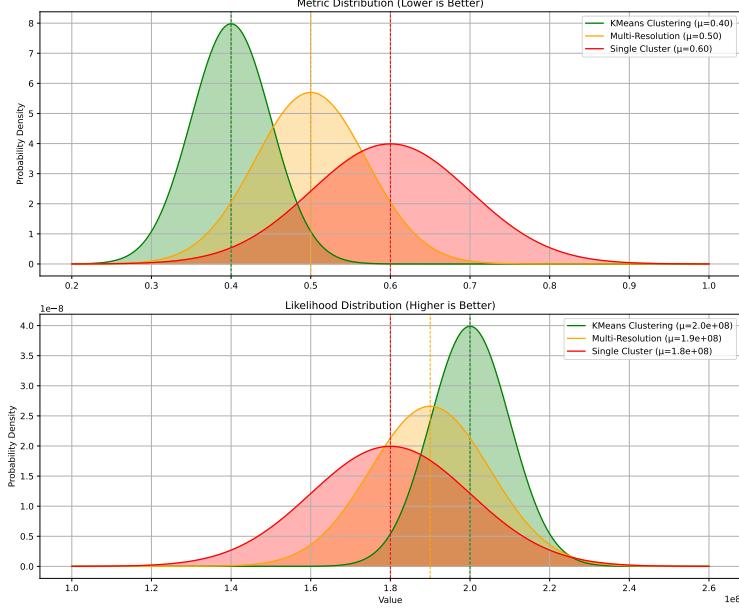


FIG. 11: NOT FINAL Distribution of CMB variance (top) and spectral negative log-likelihood (bottom) for three spatial modeling strategies: K-means clustering (green), multi-resolution grouping (orange), and single global patch (red). Dashed vertical lines show mean values. K-means clustering achieves the lowest residual variance and highest likelihood, while the single-patch model is over-constrained and the multi-resolution method shows greater statistical noise at high variance.

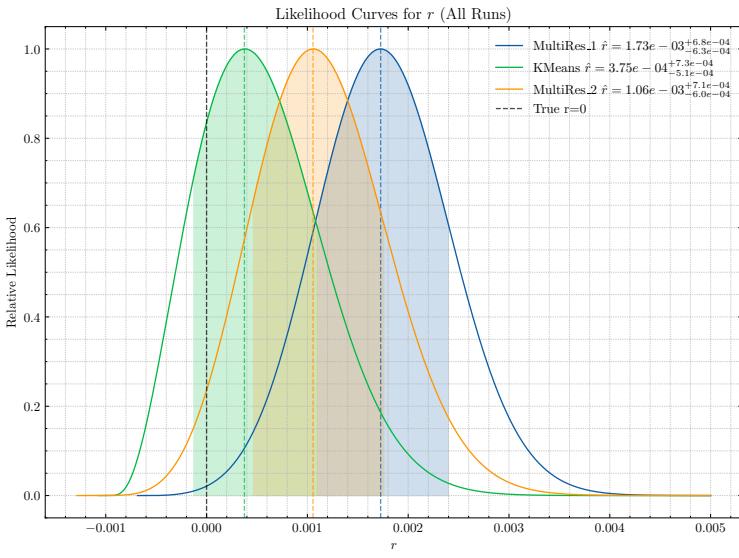


FIG. 12: Estimated tensor-to-scalar ratio likelihoods for K-means clustering (blue), multi-resolution grouping (orange), and a single global patch (green), compared to the true value $r = 0$. K-means clustering yields the lowest bias ($\hat{r} = 3.75 \times 10^{-4}$) and tightest credible interval, while MultiRes-1 and MultiRes-2 result in increasing bias and broader uncertainty as granularity increases. The single-patch model shows a strongly biased estimate with artificially narrow errors, driven by unmodeled systematic residuals. These results highlight the benefit of data-driven clustering in achieving robust and unbiased inference of cosmological parameters.

- [16] R. Stompor, S. Leach, F. Stivoli, and C. Baccigalupi, Maximum likelihood algorithm for parametric component separation in cosmic microwave background experiments, *Monthly Notices of the Royal Astronomical Society* **392**, 216–232 (2009).
- [17] J. Errard and R. Stompor, Foreground modeling and subtraction for cosmic microwave background b-mode observations, *Physical Review D* **92**, 063005 (2015).
- [18] P. Chanial, S. Biquard, and W. Kabanal, jax-healpy: Implementation of HEALPix related functions and extensions in JAX, <https://github.com/CMBSciPol/healpy>.
- [19] E. Sheldon, kmeans_radec: Kmeans clustering on the sphere, https://github.com/esheldon/kmeans_radec (2018).
- [20] Planck Collaboration, Planck 2015 results. xii. galactic astrophysics using polarized dust emission, *Astronomy & Astrophysics* **594**, A12 (2016).
- [21] K. Wassim, JAX Distributed Grid Search for Hyperparameter Tuning, <https://github.com/ASKabalan/jax-grid-search> (2025).
- [22] J. Errard and R. Stompor, Foreground modeling and subtraction for cosmic microwave background b-mode observations, *Phys. Rev. D* **92**, 063005 (2015).
- [23] J. Rader, T. Lyons, and P. Kidger, Lineax: unified linear solves and linear least-squares in jax and equinox (2023), arXiv:2311.17283 [cs.MS].
- [24] DeepMind, Optax: Gradient processing and optimization library in jax, <https://github.com/deepmind/optax> (2020).
- [25] E. Bingham, J. P. Chen, M. Jankowiak, F. Obermeyer, N. Pradhan, T. Karaletsos, D. P. Kingma, and D. Tran, NumPyro: A lightweight library for probabilistic programming in jax, <https://github.com/pyro-pp1/numpyro> (2019).
- [26] G. H. Golub and C. Reinsch, Singular value decomposition and least squares solutions (1970).
- [27] R. A. Waltz, J. L. Morales, J. Nocedal, and D. Orban, An interior algorithm for nonlinear optimization that combines line search and trust region steps, Technical Report, Northwestern University (2006), related to the TNC method used in Scipy.
- [28] R. Blondel, M. Blondel, *et al.*, jaxopt: optimization library in jax, <https://github.com/google/jaxopt> (2021).
- [29] P. Virtanen, R. Gommers, T. E. Oliphant, *et al.*, Scipy 1.0: Fundamental algorithms for scientific computing in python, *Nature Methods* **17**, 261 (2020).
- [30] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization (2017), arXiv:1412.6980 [cs.LG].
- [31] R. Stompor, J. Errard, and D. Poletti, Forecasting performance of cmb experiments in the presence of complex foreground contaminations, *Physical Review D* **94**, 10.1103/physrevd.94.083526 (2016).
- [32] J. Errard and R. Stompor, Characterizing bias on large scale cmb $\mathfrak{jmml:math}$ $\text{xmlns:mml}=\text{http://www.w3.org/1998/math/mathml"}$ $\text{display}=\text{"inline"}$ $\mathfrak{jmml:mi}, \mathfrak{bm:mi}, \mathfrak{mm:mi}, \mathfrak{mj:mi}, \mathfrak{mm:math}$ -modes after galactic foregrounds cleaning, *Physical Review D* **99**, 10.1103/physrevd.99.043529 (2019).
- [33] P.-E. G. S. Group, J. Borrill, S. E. Clark, J. Delabrouille, A. V. Frolov, S. Ghosh, B. S. Hensley, M. D. Hicks, N. Krachmalnicoff, K. Lau, M. M. Norton, C. Pryke, G. Puglisi, M. Remazeilles, E. Russier, B. Thorne, J. Yao, and A. Zonca, Full-sky models of galactic microwave emission and polarization at sub-arcminute scales for the python sky model, (2025), arXiv:2502.20452 [astro-ph.CO].
- [34] A. Zonca, B. Thorne, N. Krachmalnicoff, and J. Borrill, The Python Sky Model 3 software, *Journal of Open Source Software* **6**, 3783 (2021).
- [35] B. Thorne, J. Dunkley, D. Alonso, and S. Næss, The Python Sky Model: software for simulating the Galactic microwave sky, *Monthly Notices of the Royal Astronomical Society* **469**, 2821–2833 (2017).
- [36] A. Lewis, A. Challinor, and A. Lasenby, Efficient computation of cosmic microwave background anisotropies in closed friedmann-robertson-walker models, *The Astrophysical Journal* **538**, 473–476 (2000).
- [37] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, Optuna: A next-generation hyperparameter optimization framework, in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining* (2019) pp. 2623–2631.