The Atacama Cosmology Telescope: map-based noise simulations for DR6

The Atacama Cosmology Telescop collaboration

E-mail: zatkins@princeton.edu

Abstract. The increasing statistical power of cosmic microwave background (CMB) datasets requires a commensurate effort in understanding their noise properties. The noise in maps from ground-based instruments is dominated by large-scale correlations, which poses a modeling challenge. This paper develops novel models of the complex noise covariance structure in the Atacama Cosmology Telescope Data Release 6 (ACT DR6) maps. We first enumerate the noise properties that arise from the combination of the atmosphere and the ACT scan strategy. We then prescribe a class of Gaussian, map-based noise models, including a new wavelet-based approach that uses directional wavelet kernels for modeling correlated instrumental noise. The models are empirical, whose only inputs are a small number of independent realizations of the same region of sky. We evaluate the performance of these models against the ACT DR6 data by drawing ensembles of noise realizations. Applying these simulations to the ACT DR6 power spectrum pipeline reveals a $\sim 20\%$ excess in the covariance matrix diagonal when compared to an analytic expression that assumes noise properties are uniquely described by their power spectrum. Along with our public code, `mnms`, this work establishes a necessary element in the science pipelines of both ACT DR6 and future ground-based CMB experiments such as the Simons Observatory (SO).

Contents

1 Introduction

Increasingly precise cosmic microwave background (CMB) instruments and data products have the potential to place state-of-the-art constraints on astrophysical, cosmological, and theoretical models. Currently, ground-based instruments collect data at both small angular scales, including the Atacama Cosmology Telescope (ACT) [35, 43, 83] and the South Pole Telescope (SPT) [5, 11]; and large angular scales, including POLARBEAR [80], CLASS [24], and BICEP/Keck [49]. While these instruments observe the sky with $\mathcal{O}(10^3-10^4)$ detectors, next-generation projects, such as the Simons Observatory (SO) [37], the BICEP Array [76], and AliCPT [75] will increase this count by an order of magnitude once fully operational. Thereafter, the CMB-S4 experiment [1, 9] plans to deploy yet another order of magnitude increase in detector count, to several hundred thousand. With growing statistical power, the need for accurate modeling of instrument systematics and noise properties will become more urgent.

Accurate, computationally tractable noise models are a critical component for statistical inference with these new data. For instance, accurate noise models are necessary for an unbiased reconstruction of the CMB power spectrum covariance matrix (e.g. [32, 45, 53]); for the subtraction of dominant bias terms in CMB lensing reconstruction (e.g. [41, 50, 56]); to check for biases and quantify the variance of internal linear combination (ILC) foreground mitigation pipelines (e.g. [27, 57]); for unbiased Bayesian/parametric component separation pipelines (e.g. [28, 33, 81]); for optimal Wiener filtering or inverse-covariance filtering used in, for example, bispectrum [79] or lensing [78] estimation. A unique difficulty in the noise modeling of ground-based CMB instruments, however, is the dominance of complicated, spatially-red atmospheric noise from broad microwave water lines (e.g. [30, 61, 82]) at large scales. Importantly, increasing detector counts do not circumvent this noise component. Therefore, our experience with current CMB experiments offers a realistic preview of the noise modeling challenge facing future surveys.

Existing noise modeling frameworks either do not scale well to ground-based CMB data volumes, or simply ignore prominent features in the data. A reasonably successful, though resource-intensive, approach has been to model instrumental noise in the time-domain, and then propagate noise timestream realizations into map space (e.g. [3, 12, 20, 65]) through a mapmaking pipeline. For *Planck*, 25 million CPU hours were spent processing the FFP8

(full focal plane) simulations [69]. As the size of current and future ground-based experiments' time-ordered data (TOD) exceeds that of *Planck*,[1] this route will grow more costly. Many analyses also make use of analytic noise models from mapmaker products, such as low-resolution, dense pixel-pixel noise covariance matrices from *WMAP* [48] or per-pixel noise variance maps from *Planck* [64]. For high-resolution ground-based experiments, the former approach is numerically intractable and the latter approach neglects the strongly correlated atmospheric noise. Other analyses have taken a hybrid approach to model both spatially-varying, uncorrelated noise variance and large-scale noise correlations simultaneously (e.g. [26]). More recently, ACT [19, 55] and SPT [60] have developed noise models that are more appropriate for ground-based instruments. In particular, both collaborations used a model that captures spatially-varying, uncorrelated noise variance, as well as a stationary, two-dimensional (2D) "stripy" component (defined in either Fourier space or spherical harmonic space) resulting from the scan strategy.[2] Nevertheless, by limiting the non-white noise to a stationary representation, these models do not generalize to surveys with wide sky coverage, such as the complete ACT survey, or the SO and CMB-S4 planned surveys.

ACT's sixth data release (ACT DR6) will deliver microwave sky maps in three frequency bands (f090, f150, and f220), covering $\sim$40% of the sky to a cumulative depth of $\lesssim 15\,\mu$K-arcmin ($\lesssim 20\,\mu$K-arcmin) in temperature (polarization). In this paper, we develop novel noise models for ACT DR6, which aim to capture the observed, complicated map space noise structure inherent to ground-based CMB instruments. These models are empirical: they are built directly from ACT maps and make few *a priori* assumptions about particular map properties. We begin with an adapted version of a *tiled noise model* from [63], which builds a stationary, 2D noise model within interleaved tiles on the sky. We generalize the formalism of this model into a class of Gaussian noise models in arbitrary bases; this allows us to prescribe an *isotropic wavelet noise model* and a *directional wavelet noise model*. The latter two wavelet approaches draw on existing wavelet formalism in Fourier space (e.g. [36]) and harmonic space (e.g. [16, 52, 59, 85]), which have been applied to CMB data (e.g. [23, 27, 70, 74]), large-scale structure (e.g. [4, 17]) and Galactic foregrounds (e.g. [72]). While the isotropic model uses the wavelets of [52, 85], we also prescribe a new, tunable, directional wavelet set in Fourier space. In addition, our directional wavelets allow the lossless transform to and from wavelet space. To our knowledge, these models represent the first application of 2D wavelet analysis in the context of astrophysical instrument noise modeling. The models' primary user interface is the drawing of simulated noise maps, and we make our code — mnms[3] — publicly available. These simulations are also consumed by downstream ACT DR6 analyses, to be described in upcoming publications (e.g. [23, 71]).

This paper is structured as follows. In section 2, we summarize the ACT instrument and DR6 data products that enter the noise modeling. In section 3, we define the map-based noise and establish its most salient properties in the ACT data. In section 4, we detail the construction of the three noise models, including their mathematical formalism. In section 5, we test the noise models' ability to reproduce the complex noise covariance structure in the data, and apply them to the ACT DR6 CMB power spectrum pipeline. We discuss

---

[1]The total TOD volume included in ACT DR6 is $\sim$190 TB. For *Planck*, the total TOD volume is $\sim$13 TB. SO is expected to generate $\sim$2 TB of TOD per *day* of observations.

[2]Henceforth, we will refer to "spherical harmonic space" by a shorthand, "harmonic space". Also, while "stationary" typically describes 1D signals over time, here we take it to mean "invariant under translations in the 2D space discussed" — map space, Fourier space, or harmonic space.

[3]mnms (Map-based Noise ModelS) is available at https://github.com/simonsobs/mnms. The repository provides information on a publicly-available release of noise models and simulations. It also provides code tutorials and documentation.

implications and future directions for noise modeling, and conclude in section 6.

Unless otherwise stated, in this paper, we refer to scalar quantities by italicized variables of any case, vector quantities as boldface variables of lower case, and matrix quantities as boldface variables of upper case. We also define multiplication ($*$) and division ($/$) of vectors to signify element-wise operations.

2 The ACT instrument and products

The Atacama Cosmology Telescope (ACT) observed the polarized microwave sky from an altitude of $\sim$5,200 m in the Atacama Desert, Chile. Its receiver has gone through several upgrades since its initial deployment in 2007; this paper — and DR6 products generally — focus on data collected since its most recent upgrade, Advanced ACT, in 2016 [43].[4] In particular, we focus on data collected by three dichroic detector modules, known as "polarization arrays" (PAs, for short): PA4 (deployed 2016–2022) observing in the f150 and f220 frequency channels [46], PA5 (deployed 2017–2022) observing in the f090 and f150 frequency channels, and PA6 (deployed 2017–2019), observing the same frequency channels as PA5 [18]. Additional properties of these detector arrays are given in table 1. In 2019, PA6 was exchanged with a new detector array, PA7, observing in the f030 and f040 frequency channels. These "low frequency" measurements are not discussed in this paper. For arrays PA4–6, we use the DR6 "night only" maps, which include only data taken between 23:00 and 11:00 UTC; however, the methods we develop are equally applicable to ACT daytime data and null test maps.[5] We make use of three ACT DR6 map-based products in this paper as follows.

Sky maps. The mapmaking for DR6 will be described in a forthcoming paper; the following is a summary. The detector timestreams of each array are reduced into Stokes I, Q, and U sky maps by a maximum-likelihood mapmaking algorithm [2, 30, 55, 62]. The ACT mapmaker assumes the following data model:

$$\mathbf{d}_t = \mathbf{P}\mathbf{m} + \mathbf{n}_t \tag{2.1}$$

where $\mathbf{d}_t$ are TODs, $\mathbf{P}$ is a projection matrix mapping the pixelized sky $\mathbf{m}$ from map space to TOD space (using telescope pointing), and $\mathbf{n}_t$ is additive, Gaussian noise in each detector sample. Each data vector in equation (2.1) contains intensity and polarization components. The maximum-likelihood map under this model is:

$$\mathbf{m} = (\mathbf{P}^T\mathbf{N}_t^{-1}\mathbf{P})^{-1}\mathbf{P}^T\mathbf{N}_t^{-1}\mathbf{d}_t \tag{2.2}$$

where $\mathbf{N}_t$ is the covariance of $\mathbf{n}_t$. TODs are grouped from the same array and frequency band into the same vector $\mathbf{d}_t$, thus producing per-array, per-frequency, polarized (Stokes I, Q, U) maps. In ACT DR6, equation (2.2) is solved in eight disjoint "splits" of the timestreams: one-day-long TOD blocks are distributed exclusively into eight sets, such that the noise in each split map is independent. As in ACT DR4 [2, 58], the maps are produced in the *Plate-Carrée* cylindrical projection. Further properties of the DR6 maps are given in table 2.[6]

---

[4]ACT ended observations in late 2022.

[5]Distinguishing features of datasets such as their beam, overall depth, sky footprint, etc. make no difference to our empirical modeling. Indeed, the methods are not even restricted to maps of the CMB.

[6]This paper uses the first science-grade version of the ACT DR6 maps, labeled dr6.01. These maps are used in the upcoming ACT DR6 lensing analysis [71] and component separation analysis [23]. Since these maps were generated, the ACT collaboration has made some refinements to the mapmaking that improve the

Array	Band	Frequencies (GHz)	Beam (arcmin)
PA4	f150	124–172	1.4
	f220	182–277	1.0
PA5	f090	77–112	2.0
	f150	124–172	1.4
PA6	f090	77–112	2.0
	f150	124–172	1.4

Table 1. Properties of the Advanced ACT detector arrays used in DR6. We will refer to particular polarized maps by their array and band (e.g., PA4 f220) throughout the text. The frequency ranges contain the central 99% of the band power. Beam sizes refer to the average beam full-width at half-maximum (FWHM).

Dec., R.A. (deg)	Shape (Rows, Cols.)	Pix. Size (arcmin)	$\ell_{\max}$ (Nyq.)
−63 to 23, 180 to −180	10,320, 43,200	0.5	21,600

Table 2. Properties of the ACT DR6 maps for all arrays, bands, and Stokes polarization components. In addition to the data maps, these also apply to the inverse-variance and cross-linking maps. The $\ell_{\max}$ gives the Nyquist spherical-harmonic bandlimit of the pixelization.

Inverse-variance maps. In addition to maps of the microwave sky signal, the mapmaking algorithm produces auxiliary products used later in this paper. Of particular importance are "inverse-variance" maps, which give the *inverse* of the white-noise covariance between Stokes components in each pixel. Following the ACT convention [19, 57], we label them **h**. The inverse-variance maps are analogous to detector hit-count maps, but also include the effect of intrinsic detector properties per detector hit [58]. Because the TOD noise power spectrum is red, with a "white-noise floor" at high frequencies [30, 61], the inverse-variance maps are only accurate at small scales. Although the Stokes covariance in each pixel is formally a 3×3 matrix, for the same reasons as in DR4 (see [2]), DR6 includes only the Stokes intensity-intensity (II) element [2]. This has little effect on our noise models, however. As we will show in section 4, our models still account for correlated Stokes components in the DR6 maps. An example inverse-variance map is shown in the top panel of figure 1.

Cross-linking maps. The mapmaker also produces "cross-linking maps", which we use as a reference in section 3.2 and to guide our modeling decisions in section 4. These are maps which contain information about the ACT receiver's local scanning direction (see e.g., [2, 19, 58], for a more complete description). In particular, the "polarization fraction", $\mathbf{f}_p$, of the cross-linking map encodes the degree to which a given pixel is observed by many different scanning directions. The cross-linking maps are produced in such a way that $\mathbf{f}_p = 0$ corresponds to maximal cross-linking (many crossing scans), and $\mathbf{f}_p = 1$ corresponds to no cross-linking (all scans in the same direction). Therefore, we use the quantity $1 - \mathbf{f}_p$ in later sections, such that high values of $1 - \mathbf{f}_p$ correspond to high cross-linking and vice versa. An
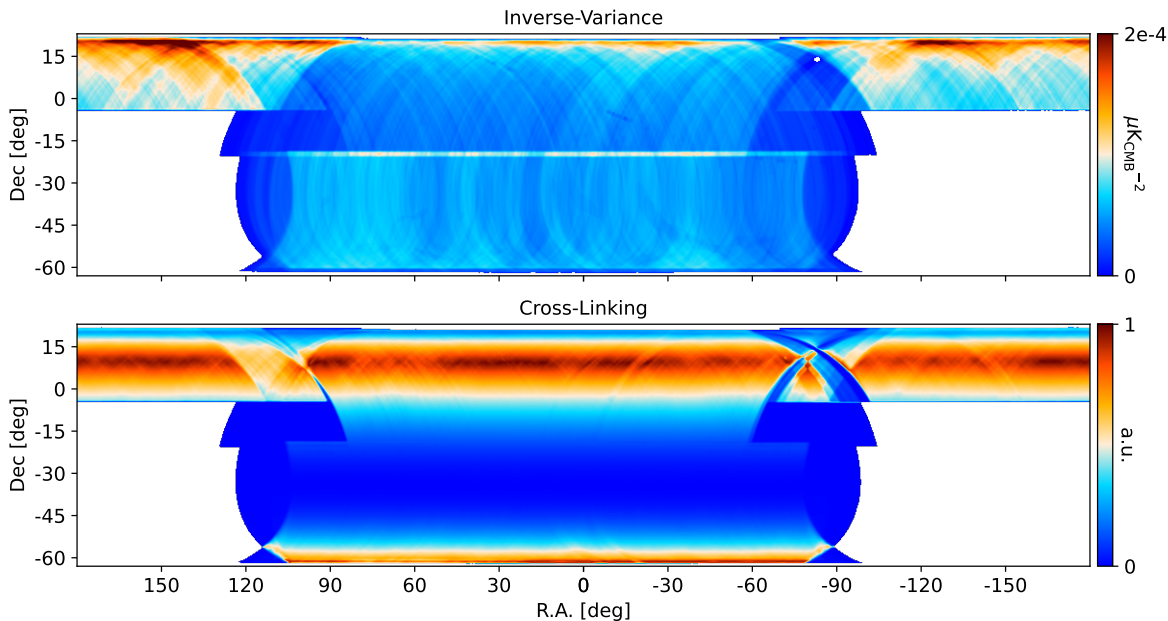
Figure 1. *Top*: the inverse-variance map, $\mathbf{h}$, for PA5 f090, first split. Larger values correspond to deeper regions of the map. The ACT scanning strategy is also discernible. Border ticks give the declination (Dec) and right-ascension (R.A.) in degrees. *Bottom*: the cross-linking map, $1 - \mathbf{f}_p$, for the PA5 f090 coadd (see equation (3.2)). Values close to 0 indicate poor cross-linking, and vice versa for values close to 1, in arbitrary units (henceforth, a.u.).

example cross-linking map is shown in the bottom panel of figure 1. Both the inverse-variance and cross-linking maps are produced with the same projection and parameters (see table 2) as the sky maps.

3 Map-based noise in ACT data

We begin our analysis by exploring the map-based noise in the ACT dataset. To do so, we first define "map-based noise" and prescribe how we measure it.

3.1 Definition and measurement

Each pixelized, polarized (Stokes I, Q and U) split map has an independent noise realization:

$$\mathbf{m}_i \equiv \mathbf{s} + \mathbf{n}_i \tag{3.1}$$

where $\mathbf{m}_i$ is the map solution in TOD split i, $\mathbf{s}$ is the underlying sky signal (assumed constant over splits), and $\mathbf{n}_i$ is additive noise in each pixel (we suppress indices labeling the array, frequency, and polarization). The form of equation (3.1) is not arbitrary — it follows from the additive time-domain noise in the definition of $\mathbf{d}_t$ (equation (2.1)) and the linearity of the mapmaking solution (equation (2.2)). That is, $\mathbf{n}_i$ is the mapped time-domain noise.

This paper aims to model the noise, $\mathbf{n}_i$, in each ACT split. Of course, we can never measure $\mathbf{n}_i$ directly, or else we could produce noiseless maps. Rather, we must use a proxy for the noise. In particular, we use differences of map splits, relying on the assumption that each $\mathbf{m}_i$ in equation (3.1) measures an identical signal. We give the prescription by which we construct these difference maps as follows.

The first step is to construct an unbiased estimate of the signal, $\mathbf{s}$, in the split maps. We do this by taking the inverse-variance-weighted average of $\mathbf{m}_i$ over splits i, which we call the map-based "coadd", $\mathbf{c}$:

$$\mathbf{c} = \frac{\sum_i \mathbf{h}_i * \mathbf{m}_i}{\boldsymbol{\Sigma}} \tag{3.2}$$

where $\mathbf{h}_i$ is the inverse-variance map for split i and $\boldsymbol{\Sigma} \equiv \sum_i \mathbf{h}_i$. Next, we subtract $\mathbf{c}$ from each of the $\mathbf{m}_i$ to construct a "raw" difference map, $\mathbf{d}_i$:

$$\mathbf{d}_i \equiv \mathbf{m}_i - \mathbf{c}. \tag{3.3}$$

If $\mathbf{c}$ were exactly $\mathbf{s}$, then $\mathbf{d}_i$ would equal $\mathbf{n}_i$ in that split. Instead, because $\mathbf{c}$ contains an average of the noise from each split, so too does $\mathbf{d}_i$. We can see this by combining equations (3.1), (3.2), and (3.3):

$$\mathbf{d}_i = \mathbf{n}_i - \frac{\sum_j \mathbf{h}_j * \mathbf{n}_j}{\boldsymbol{\Sigma}} \tag{3.4}$$

The upshot, though, is that $\mathbf{d}_i$ is a decent proxy for $\mathbf{n}_i$: it contains *only* noise, and "mostly" (which we will quantify in equation (3.8)) noise from split i. Indeed, this is the proxy used in previous ACT analyses [19, 57, 63].

In this paper, we perform one more step by applying a correction factor to the raw difference maps. Specifically, we correct for the fact that the variance in $\mathbf{d}_i$ is less than the variance in $\mathbf{n}_i$. To evaluate $\mathrm{Var}(\mathbf{d}_i)$, first recall from section 2 that $\mathrm{Var}(\mathbf{n}_i) = 1/\mathbf{h}_i$ only in the white noise, uncorrelated-pixel limit. We work in that limit for now, and also define the difference map inverse-variance along the same lines: $\mathrm{Var}(\mathbf{d}_i) \equiv 1/\mathbf{w}_i$. In appendix C, we show that $\mathbf{w}_i$ satisfies:

$$\frac{1}{\mathbf{w}_i} = \frac{1}{\mathbf{h}_i} - \frac{1}{\boldsymbol{\Sigma}}. \tag{3.5}$$

That is, the left-hand side — the variance of $\mathbf{d}_i$ — is less than that of $\mathbf{n}_i$ by exactly $1/\boldsymbol{\Sigma}$. This motivates the following correction to the raw difference maps:

$$\boldsymbol{\nu}_i \equiv \mathbf{d}_i * \sqrt{\frac{\mathbf{w}_i}{\mathbf{h}_i}} \tag{3.6}$$

such that the variance of $\boldsymbol{\nu}_i$ is the same as that of $\mathbf{n}_i$:

$$\mathrm{Var}(\boldsymbol{\nu}_i) = \mathrm{Var}(\mathbf{d}_i) * \frac{\mathbf{w}_i}{\mathbf{h}_i} = \frac{1}{\mathbf{w}_i} * \frac{\mathbf{w}_i}{\mathbf{h}_i} = \frac{1}{\mathbf{h}_i} = \mathrm{Var}(\mathbf{n}_i). \tag{3.7}$$

Both when we examine map-based noise properties of a given split (section 3.2), and build noise models (section 4), we will always be referring to the *corrected* difference map noise measurements, $\boldsymbol{\nu}_i$, in equation (3.6).

We emphasize that $\boldsymbol{\nu}_i$ and $\mathbf{n}_i$ are distinct but related quantities. Their correlation is given by (appendix C):

$$\boldsymbol{\rho}_i = \sqrt{1 - \frac{\mathbf{h}_i}{\boldsymbol{\Sigma}}} \tag{3.8}$$

which approaches 1 as the number of splits increases, assuming splits contain reasonably uniform TOD volumes. For example, given eight approximately uniform splits, equation (3.8) yields a correlation of $\sim 93.5\%$. Because we can never directly measure $\mathbf{n}_i$ in a map, we resort to using $\boldsymbol{\nu}_i$ as a highly correlated tracer.

Fortunately, we can test the two assumptions that the uncorrected difference maps (equation (3.3)) contain pure noise, and that the corrected difference maps (equation (3.6)) have accurate variance. From the best-fit model of the signal from ACT DR4 [19], the expected BB power from the lensed CMB and Galactic dust (with the Galaxy masked, as in this paper's pseudospectrum mask, see section 4.5.3) should be at the sub-percent level compared to the ACT DR6 per-split noise. In other words, the BB power spectra from raw, *non-differenced* split maps, $\mathbf{m}_i$, should be a reasonably clean measure of the map-based noise; in particular, they should match BB power spectra from the corrected-power difference maps, $\boldsymbol{\nu}_i$. Such a test is not trivial: *Planck* [69] found a several percent discrepancy in their High Frequency Instrument (HFI) channels when performing a similar comparison, which was attributed to systematics in their pre-processing.

As shown in figure 2, ACT DR6 exhibits good agreement. All arrays and frequencies achieve an average difference map ($\boldsymbol{\nu}_i$) to raw map ($\mathbf{m}_i$) BB spectra ratio of within 0.1% of nominal. Recall, in constructing the difference maps (equation (3.6)), we had to work in the white noise, uncorrelated-pixel limit where the inverse-variance maps, $\mathbf{h}_i$, are expected to be accurate. As discussed more in section 3.2, this limit is reached only on small angular scales. However, the performance of the $\sqrt{\mathbf{w}_i/\mathbf{h}_i}$ correction factor is more robust: the BB ratios are consistent with nominal at the $\sim 1\%$ level for scales as large as $\ell \sim 300$. While the observed percent-level deviations for $\ell \lesssim 300$ do matter when evaluating our noise models in section 5, overall this result strongly supports the scheme by which we measure the noise in a given map split.

3.2 ACT map-based noise properties

We motivate our noise model design choices by enumerating five salient features of the ACT map-based noise. We take these five features to define the scope of our modeling in section 4; however, we do not expect this list to be exhaustive in reality. Nevertheless, we do expect the noise properties in this section to be (a) the most significant for downstream analyses, and (b) common to ground-based, scanning microwave telescopes.

3.2.1 Large-scale noise correlations

Large-scale noise correlations are a dominant noise feature in ground-based microwave observatories [30, 44]. Their origin is typically attributed to fluctuating, nearly unpolarized emission by atmospheric water vapor [34, 61], which can leak into polarized timestreams via optical and detector systematics [82]. We see their effect in the power spectra of the DR6 difference maps, as shown in the top panel of figure 3. Both the intensity and polarized map-based noise power spectra are steep and red, spanning several orders of magnitude in power from their white-noise level at small scales to their peaks at large scales. The correlated noise in intensity maps is higher than in polarization, mainly due to a smaller transition scale — or, higher ℓ_{knee} — from red, correlated noise to white noise. The power-law index of the correlated noise — the slope in log-log space, as plotted in the top panel of figure 3 — is similar between intensity and polarization. As is expected for thermal water emission, the intensity spectra increase monotonically with the frequency band of the map. This scaling is not as clear in the polarization spectra, highlighting the contribution from non-atmospheric effects
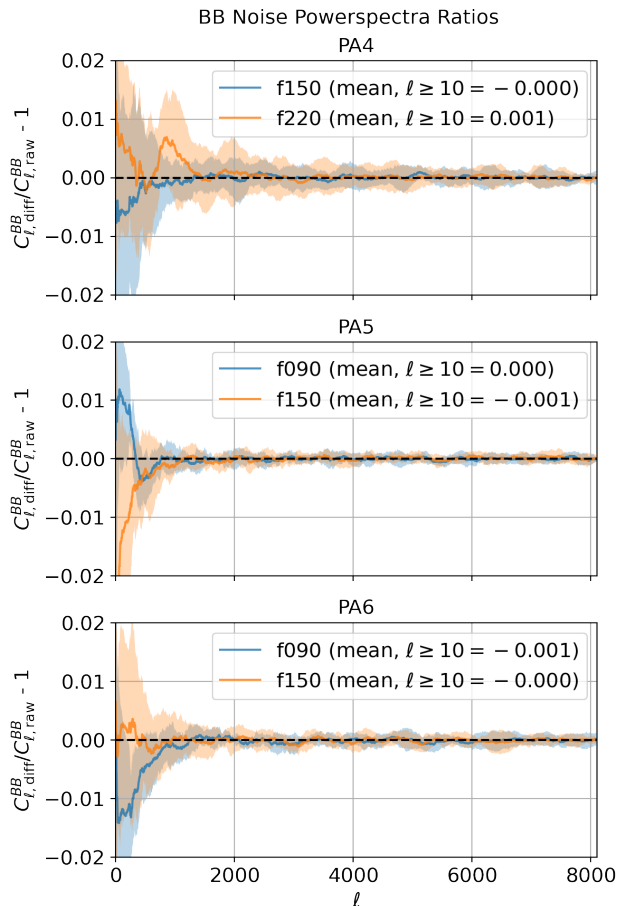
Figure 2. Ratios of BB pseudospectra between the corrected difference maps, $\boldsymbol{\nu}_i$, and the individual split maps, $\mathbf{m}_i$, for each array and frequency channel in ACT DR6. All pseudospectra are measured within this paper's "pseudospectrum mask" described in section 4.5.3. The dashed black line is the nominal result (ratio of unity), the solid lines (shaded bands) are the means (1σ range) over the eight splits of each dataset. To mitigate any expected BB signal, in addition to masking the Galaxy, all maps are point-source subtracted and filtered using the ACT DR4 ground pickup filter [19], where we remove Fourier modes with $|k_x| \leq 90$ and $|k_y| \leq 50$. Spectra are smoothed with a tophat kernel of size $\Delta\ell = 250$ to better visualize their overall trend.

to the large-scale polarized noise. Likewise, the convergence of the PA4 f220 TT spectrum to the f150 spectra at low-ℓ requires more work to understand its origin.

3.2.2 Correlated frequencies

Because the ACT detector arrays are dichroic, maps of each frequency band on a given detector array observe correlated fluctuations. We probe this by forming "correlation spectra" of the two frequency maps:

$$r_\ell^{ab} \equiv \frac{C_\ell^{ab}}{\sqrt{C_\ell^{aa} C_\ell^{bb}}} \tag{3.9}$$

where C^{ab} denotes the cross-spectrum of two fields a and b. The bottom panel of figure 3 shows how the correlation spectra vary by detector array, polarization, and angular scale.
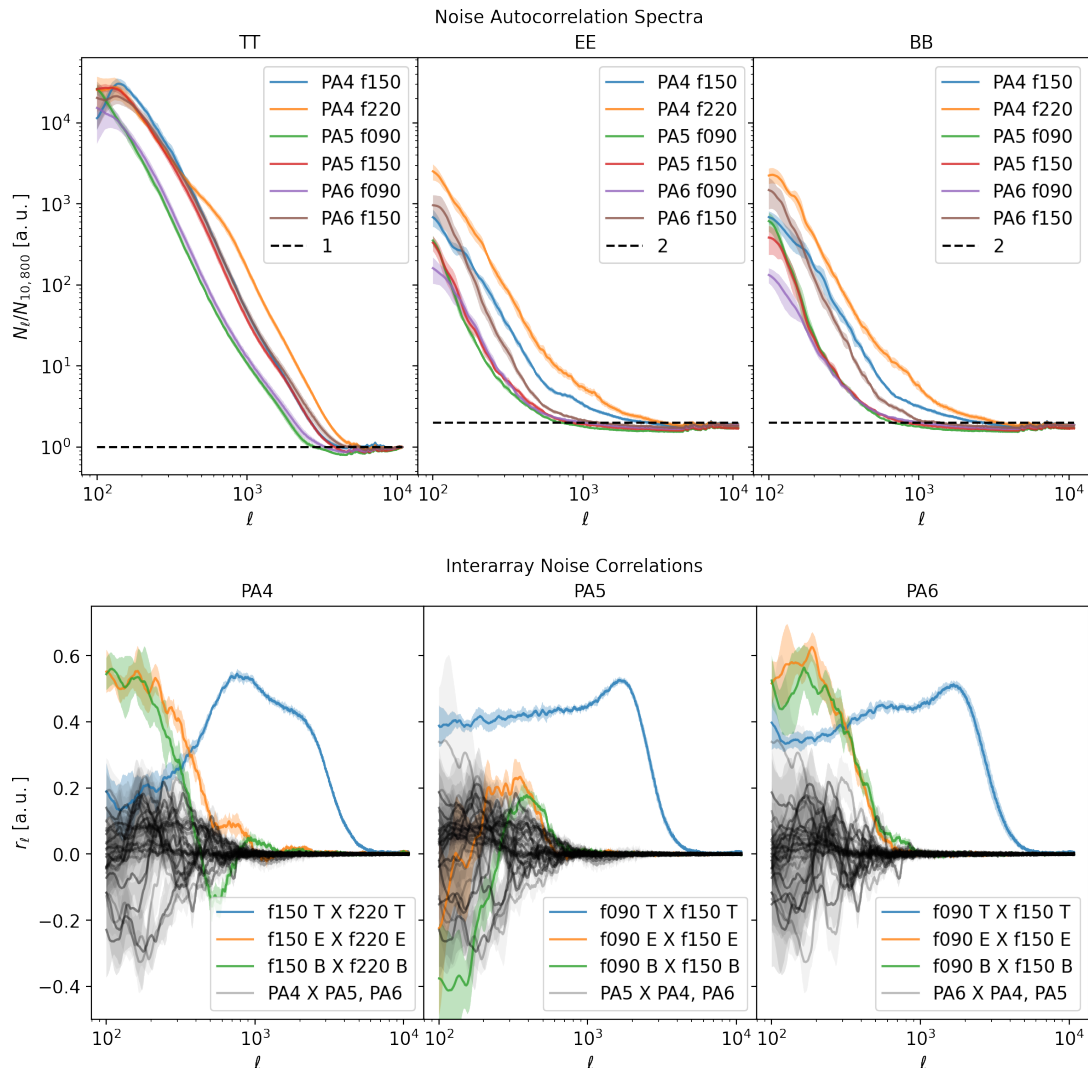
Figure 3. *Top, large-scale noise correlations*: noise power spectra by polarization autocorrelation (TT, EE, BB), for each detector array and frequency band. The power spectra are evaluated on the corrected difference maps (equation (3.6)), after applying the apodized pseudospectrum mask described in section 4.5.3. Shaded bands give the 1σ range over the eight split maps for each array/frequency. All spectra are smoothed with a tophat kernel of size $\Delta\ell = 25$. Both the TT and polarization (EE, BB) spectra are normalized by the value of the TT spectrum at $\ell = 10,800$ to suppress the effect of variable data volume between arrays. Polarization spectra plateau at a value of ~ 2, owing to each polarized component receiving approximately half the TOD volume as intensity. Both the intensity and polarization spectra are steep and red at large scales. *Bottom, correlated frequencies*: correlation spectra (r_ℓ, equation (3.9)) evaluated from the mean cross-power spectrum of the eight difference maps, for each detector array. The same mask is applied as in the top panel, and shaded regions again correspond to the 1σ range over the eight split maps. For legibility, we only plot the cross-spectra between different frequency bands — all crosses are between the same polarization components (i.e., TT, EE, or BB). The colored series indicate cross-spectra between frequency bands on the same physical detector array, while the grey series indicate the cross-spectra between bands on different physical detector arrays. The only significant cross-spectra come from the same physical detector array.

The TT correlation spectra between frequency maps on the *same* array stand out in particular: the two frequencies are $\sim$25%–50% correlated at scales where the atmospheric noise is dominant ($\ell \lesssim 3,000$, see the top panel figure 3). Likewise, the polarization (EE, BB) correlation spectra between frequency maps, again on the *same* array, appear significant to similar levels on larger scales (though more for PA4 and PA6 than PA5). Owing to the increased distance between detectors, correlations of any component between frequency maps on *different* physical detector arrays are considerably reduced. Other details apparent in figure 3 are less intuitive. It is not clear why the TT correlation peaks at $\ell \approx 1,000$ but declines for larger scales especially in PA4, nor why PA5 polarization correlations are considerably different from those of PA4 and PA6. Further work is needed to understand these features. Nevertheless, our models treat these empirical noise correlations in the ACT maps as a given.

3.2.3 Map-depth inhomogeneity

Map-depth inhomogeneity[7] is a commonly modeled property of map-based noise in the literature. For ACT, this information is encoded in the inverse-variance maps $\mathbf{h}$ discussed in section 2. The inverse-variance maps contain considerable morphology, as is seen in the top panel of figure 1. In particular, we see map depth varies spatially on both large and small scales, generally being deeper closer to the celestial equator and shallowest near the Galactic plane. As discussed, we expect this spatial modulation to only be accurate for the high-ℓ ACT noise. Furthermore, we can discern the telescope scan strategy inducing coherent arcs across the observed footprint. Between declinations of $\sim 0°$ to $15°$, the ACT scan strategy tends to have an "x-like" directional pattern, while declinations of $\sim -45°$ to $-15°$ have a uniformly "vertical" scanning pattern. The scanning pattern is reflected in the bottom panel of figure 1, which contains a cross-linking map, $1 - \mathbf{f}_p$ (see section 2). Regions where the cross-linking is low are those where the ACT scans are uniformly vertical; regions where the cross-linking is high are those where the ACT scans are "x-like".

3.2.4 Spatially-varying noise anisotropy

Unfiltered ACT data have historically shown stripy noise features [19, 55], and the DR6 maps are no exception. Like the large-scale correlations (figure 3), stripy noise patterns are a manifestation of the atmospheric noise. Given a particular low-frequency atmospheric signal in a TOD, as the telescope scans across the sky, the signal will manifest as a one-dimensional stripe in the map. Thus, not only are the largest spatial scales dominated by atmospheric signals, but those signals also tend to "trace" the ACT scan pattern. We confirm this intuition by examining the 2D Fourier noise power in the previously mentioned declination bands, as shown in figure 4. In order to emphasize the "anisotropic" noise power, we first divide-out the "isotropic" radial profile (i.e., the 1D spectra from figure 3) that would otherwise dominate the full 2D power. Near the celestial equator (the left panel), we recover a strong "x-like" pattern in the relative 2D noise power, as expected from our by-eye assessment of the ACT scan pattern. Likewise, in the southern region with vertical, poorly cross-linked scans (the right panel), we find the 2D noise power to be concentrated in a horizontal "bar" in Fourier space. In other words, the 2D noise power exhibits long-wavelength (small wavenumber) correlations in the y-direction, but is approximately featureless (white) in the x-direction, as expected.

---

[7]Herafter, we take "inhomogeneities" to mean quantities that vary over 2D position on the sky, and "anisotropies" to mean quantities that vary over "azimuthal direction" in the maps, as centered on any given 2D position. These differ from their conventional definitions in cosmology.
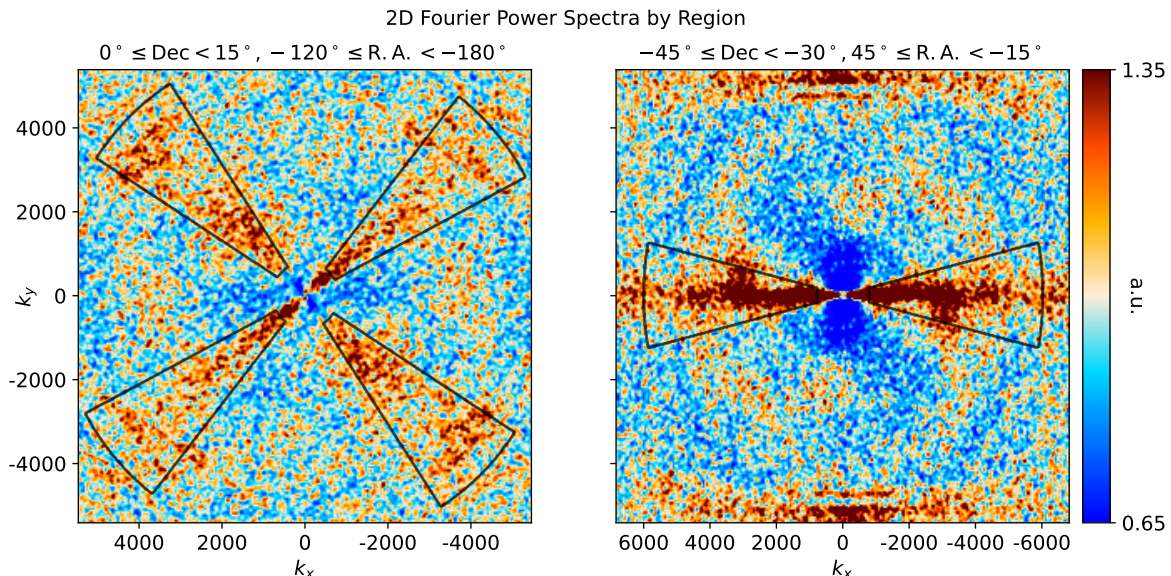
Figure 4. *Spatially-varying noise anisotropy*: TT power spectra in 2D Fourier space for PA5 f090, first split. The 2D spectra are normalized radially by average power in annuli of $\Delta k = 50$ and smoothed by a uniform 2D tophat kernel of sidelength $\Delta k = 100$. The left spectrum is measured in the region from $0°$ to $15°$ Dec and $-120°$ to $-180°$ R.A.. The right spectrum is measured in the region from $-45°$ to $-30°$ Dec and $45°$ to $-15°$ R.A.. The edge of both regions is apodized with a $2°$-wide cosine taper before taking Fourier transforms. Coordinates at the image perimeter give the vertical and horizontal Fourier modes. Because the 2D spectra are normalized by their overall radial profile, the colorscale is unitless. As is explained in the text, the noise exhibits stripy patterns that align with the local scan strategy. In Fourier space, these patterns manifest as bars perpendicular to the local scan strategy. We outline the primary features from the scanning strategy in black. The left (right) panel outlined region contains an average power excess of 8% (25%). Secondary features — such as detector-detector correlations on the top and bottom of the right panel — are also visible. The regions of low noise power near the center result from the removal of the radial, or "isotropic", part of the power spectrum.

Furthermore, the poorly cross-linked region exhibits stronger overall anisotropies than the well cross-linked region. To demonstrate this, in figure 4, we measure the power excess within the dominant "x-like" or "bar" features in the 2D spectra. The outlined "x-like" portion of the well cross-linked 2D spectrum contains an average noise power excess of 8%, whereas the same as measured in the outlined "bar" of the poorly cross-linked 2D spectrum is 25%. Thus, the ACT map-based noise contains anisotropies whose structure and prominence varies spatially.

3.2.5 Scale-dependent map depth

As discussed above, the ACT map-based noise is spatially modulated by the inverse-variance maps at small scales. At large scales where the atmospheric correlations dominate, however, the map depth spatial morphology resembles the amount of cross-linking.[8] This scale-dependent behavior can be understood as a consequence of the ACT mapmaking solution itself. At small scales where the noise is uncorrelated between detectors, variance in the maximum likelihood map in equation (2.2) decreases only with the number of detector hits. This

---

[8]An equivalent description of scale-dependent map depth is a spatially-dependent noise power spectrum.

does not apply to large-scale correlated modes observed simultaneously by many detectors. Instead, variance in the estimation of large scales decreases with the number of independent scans that observe the mode (heuristically, for atmospheric correlations on the order of the array size, the array acts as if it has only one detector, not hundreds). All scans are not equally effective at reducing variance in the estimate, however. In a single scan, large-scale, correlated signals on the sky are degenerate in equation (2.2) with low-frequency, correlated atmosphere in the TOD. Observations of a fixed sky position with differing scan directions act to lift this degeneracy, while parallel scans do not. Therefore, in addition to averaging down with the number of independent scans, the large-scale map-based noise averages down with increased map *cross-linking* as well.

To isolate this cross-linking-like morphology, we should first normalize the difference maps by the number of array scans at each point. Maps of these statistics are not readily available, so we use the inverse-variance maps as a proxy.[9] To be precise, we first multiply the difference maps, $\boldsymbol{\nu}_i$, by $\sqrt{\mathbf{h}_i}$, such that the variance of $\boldsymbol{\nu}_i$ on small-scales is uniform. Borrowing from equation (3.7), we have *on small-scales*:

$$\mathrm{Var}(\boldsymbol{\nu}_i\sqrt{\mathbf{h}_i}) = \mathrm{Var}(\boldsymbol{\nu}_i) * \mathbf{h}_i = \frac{1}{\mathbf{h}_i} * \mathbf{h}_i = 1. \qquad (3.10)$$

Then we filter $\boldsymbol{\nu}_i\sqrt{\mathbf{h}_i}$ for *large-scales* and look for residual structure in the noise power across the map. Figure 5 presents the small- and large-scale behavior of the data: the top panel shows the *inverse* of the variance of $\boldsymbol{\nu}_i$ after filtering for small scales, while the bottom panel shows the *inverse* of the variance of $\boldsymbol{\nu}_i\sqrt{\mathbf{h}_i}$ after filtering for large scales as just discussed. Comparing to figure 1, the empirical small scale power clearly traces the mapmaker-produced inverse-variance map, while the empirical large-scale power resembles the mapmaker-produced cross-linking map — with greater noise power (less inverse-variance) corresponding to less cross-linking, and vice versa. In sum, different angular scales have their own distinct "inverse-variance maps", resulting directly from the ACT scanning strategy.

Note that all of the above noise properties are each manifestations of the same underlying process: the propagation of correlated time-domain noise and the ACT scanning strategy through the mapmaking. Further study of atmospheric emission [61] and instrumental noise sources [30] may reveal new, subtle map-based noise properties not explored in this paper.

4 Noise models

In this section we describe three different noise models that capture the ACT map-based noise properties from section 3.2. Implicit in equation (3.1) is the assumption that the particular map based noise realizations in our data, $\mathbf{n}_i$, are zero-mean random vectors. We also make the assumption that the $\mathbf{n}_i$ are Gaussian distributed.[10] Therefore, a statistically complete "noise model" for each $\mathbf{n}_i$ is given by its covariance matrix, $\mathbf{N}_i$. We build estimates of this matrix using the difference maps, $\boldsymbol{\nu}_i$, and use it to draw noise simulations.

The long-distance noise correlations from the previous section suggest that $\mathbf{N}_i$ is dense in the map basis. Combined with the large pixel count in the ACT DR6 maps ($N_{\mathrm{pix}} =$

---

[9]This essentially assumes that the per-pixel number of array scans is proportional to the per-pixel number of detector hits, which is not exact but more than good enough for our purposes here.

[10]This is well motivated because the time-domain noise is already well described by Gaussian noise. The noise properties stay Gaussian when the linear mapmaking equation (equation (2.2)) is applied. Furthermore, each pixel is built up from many independent time samples ($\sim 97\%/99\%/98\%$ of pixels have greater than 100 time sample hits in each split map of PA4/5/6) which is expected to further reduce any non-Gaussianity.
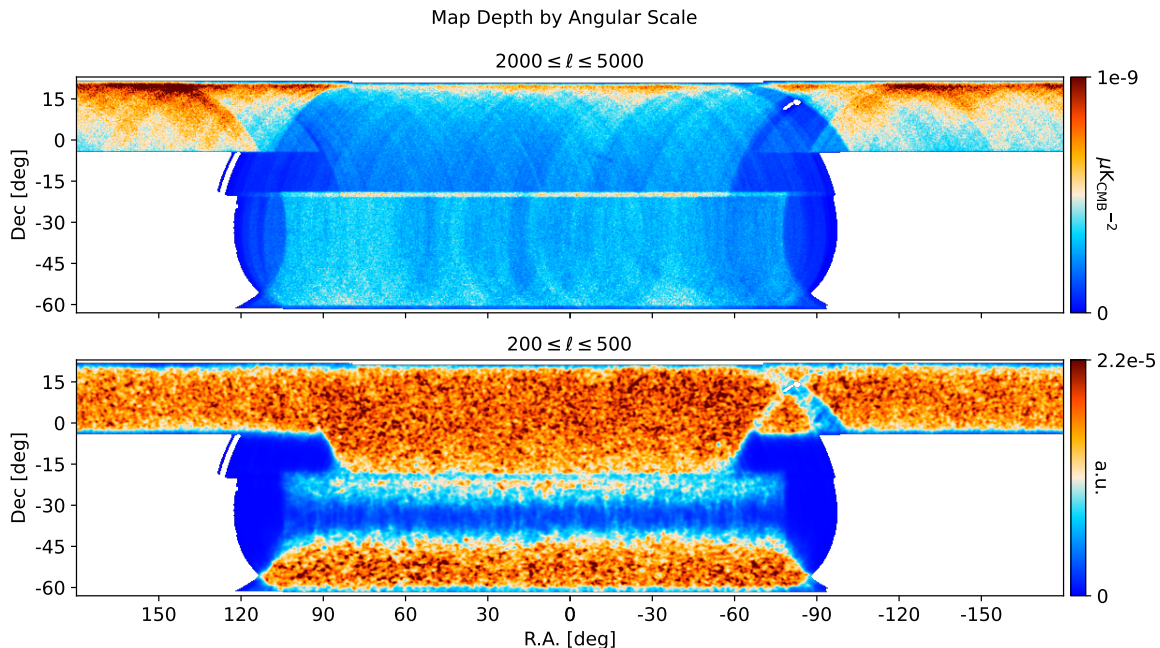
Figure 5. *Scale-dependent map depth*: "Inverse-variance maps" as measured directly from the PA5 f090 difference maps. *Top*: after filtering $\boldsymbol{\nu}_i$ for small scales, $2000 \leq \ell \leq 5000$, and averaging over Stokes Q and U in only the first split. *Bottom*: after filtering $\boldsymbol{\nu}_i \sqrt{\mathbf{h}_i}$ for large scales, $200 \leq \ell \leq 500$, and averaging over Stokes Q and U in all eight splits. We square the filtered maps to assess local noise variance, smooth them for legibility (Gaussian FWHM $= 4\pi/5000\,\mathrm{rad} \approx 0.14^o$ and FWHM $= 2\pi/500\,\mathrm{rad} \approx 0.72^o$, respectively), and take their inverse. We also multiply by pixel area to account for declination-dependent loss of power from the filtering. Because we invert the variance, small values indicate *larger* noise power. As is explained in the text, the small-scale noise power is modulated by the mapmaker inverse-variance, and the large-scale noise power traces the mapmaker cross-linking.

$445, 824, 000$, see table 2), the construction of an explicit $N_{\mathrm{pix}} \times N_{\mathrm{pix}}$ $\mathbf{N}_i$ is intractable [63]. Instead, we leverage the statistical properties in section 3.2 to build a sparse, but still approximately complete, covariance matrix. The key to a sparse representation is the isolation of stationary regions in some projection of the noise. For example, given map space fields a and b at locations $\mathbf{x}$ and $\mathbf{x}'$, and their Fourier transforms $\tilde{a}$ and $\tilde{b}$ at modes $\mathbf{k}$ and $\mathbf{k}'$, the cross-correlation theorem states:

$$
\begin{aligned}
\text{If}: \quad & \langle a(\mathbf{x})b^*(\mathbf{x}')\rangle \equiv C^{ab}(\mathbf{x}, \mathbf{x}') = C^{ab}(\mathbf{x} - \mathbf{x}') \\
\text{then}: \quad & \langle \tilde{a}(\mathbf{k})\tilde{b}^*(\mathbf{k}')\rangle \equiv \tilde{C}^{ab}(\mathbf{k}, \mathbf{k}') = \tilde{C}^{ab}(\mathbf{k})\delta(\mathbf{k} - \mathbf{k}')
\end{aligned}
\tag{4.1}
$$

where C^{ab} is the map space covariance of a and b, and $\tilde{C}^{ab}$ is its Fourier transform.[11] In other words, if the covariance of two fields is *stationary* in map space — that is, exhibits translational invariance (even while being correlated) — then in Fourier space it is *uncorrelated*. In particular, only the diagonal of the Fourier covariance — the (cross) power spectrum $\tilde{C}^{ab}$ — needs to be measured. Therefore, for each noise model discussed in this section, we will first filter the ACT DR6 noise maps to promote their stationarity, and then transform them to permit a sparse representation of the covariance matrix.

---

[11]Due to the unitary property of the Fourier transform, equation (4.1) is equally valid if we exchange map space and Fourier space.

Following a brief discussion of some additional notation in section 4.1, we begin in section 4.2 with an updated version of the "tiled" noise model from [63], including the algorithms by which we build the model and draw simulations. We describe the same for two novel noise models: an "isotropic wavelet" (or "wavelet") noise model in section 4.3 and a "directional wavelet" (or "directional") noise model in section 4.4. Throughout sections 4.2, 4.3, and 4.4, we will introduce proper matrix notation for all operations performed on our map vectors. This will aid users incorporating the noise covariance matrices into their analysis program, as well as clarify how we draw noise simulations. Lastly, for all three models, we describe map preprocessing steps and additional implementation details in section 4.5.

4.1 Linear algebra notation

Before diving in, we establish some notation accompanying our map vectors and matrices. So far, we have understood our proxy noise maps, $\boldsymbol{\nu}_i$, to have N_{pix} elements. Henceforth, we will add an additional index to the maps which iterates over frequency bands (2) and polarization indices (3) on a given detector array. That is, we should now take the vector $\boldsymbol{\nu}_i$ to refer to a concatenation of the (6) polarized frequency maps from each array, with $6N_{\mathrm{pix}}$ elements. In this way, our noise models will be able to capture the observed frequency-frequency cross-correlations from figure 3.[12] When needed, we will refer to a particular frequency and polarization subset of the data by its superscript index, a, ranging from 0 to 5 (e.g., for PA4, the $a = 0 - 2$ blocks of $\boldsymbol{\nu}_i^a$ would refer to the PA4 f150 band, I, Q, and U, respectively). This notation extends to other map vectors like $\mathbf{h}_i$, and any matrix objects we construct from the outer-product of two vectors — when needed, we will refer to matrix block elements using superscript indices a and b. Each map vector and noise model will continue to be distinct for each detector array and split, i.

We also define matrices for the map transforms — discrete Fourier transforms (DFTs) and spherical harmonic transforms (SHTs) — used in our modeling and simulation algorithms. We denote the forward DFT by the matrix $\mathbf{F}$, which is unitary ($\mathbf{F}^{-1} = \mathbf{F}^{\dagger}$). Following [73, 77], we denote spherical harmonic synthesis (from harmonic to map space) by the $N_{\mathrm{pix}} \times N_{a_{lm}}$ matrix $\mathbf{Y}$, and spherical harmonic analysis (from map to harmonic space) by $\mathbf{Y}^{\dagger}\mathbf{W}$, where $\mathbf{W}$ is a square diagonal matrix containing per-pixel "quadrature weights" used in the numerical integration. In general, it is the case that $N_{\mathrm{pix}} \neq N_{a_{lm}}$: SHTs are, unlike DFTs, rectangular operations. Unless otherwise stated, we always perform SHTs to the Nyquist bandlimit of the input pixelization. Following the default SHT implementation in `pixell`[13] and `healpy`[14] [40, 87], when applied to a spin-2 pair of Stokes Q and U maps, we will take $\mathbf{Y}^{\dagger}\mathbf{W}$ to be the linear combination of spin-2 SHTs that outputs E- and B-mode harmonic coefficients [86].[15] Thus, if $\mathbf{s}$ is a polarized map with Stokes I, Q, and U components, the vector $\mathbf{a} = \mathbf{Y}^{\dagger}\mathbf{W}\mathbf{s}$ consists of I (also referred to as T), E, and B harmonic coefficients.

Finally, we make use of a compact notation to describe products of matrices and vectors with incompatible shapes. The notation is similar to the "broadcasting" of vector and matrix objects used in `numpy`, the main scientific computing package for the Python programming language [42]. The use of this notation helps save ink, prevents redundant definitions of the same linear operation, and directly corresponds to the implementation of these operations

---

[12]While we do not observe significant polarization-polarization noise cross-correlations in DR6, in keeping with previous ACT convention [19, 57], we allow for them in the models.

[13]https://github.com/simonsobs/pixell.

[14]https://github.com/healpy/healpy.

[15]Unlike cosmological modeling, the noise modeling done in this work does not obviously benefit from a description in terms of the parity even and odd E and B fields. It is likely that a reformulation in terms of spin-0 SHTs of Q and U would have worked equally well.

in code. For example, under this notation it makes equal sense to write $\mathbf{Fm}$ for the DFT of an intensity-only map $\mathbf{m}$ as it does $\mathbf{F}\boldsymbol{\nu}_i$, where, as discussed, $\boldsymbol{\nu}_i$ is a six-component map. In the latter case, the DFT matrix $\mathbf{F}$ is understood to be promoted to a 6×6 square block-diagonal matrix with the original $\mathbf{F}$ along the block-diagonal. Likewise, $\mathbf{a} = \mathbf{Y}^\dagger \mathbf{Wm}$ is a spin-0 harmonic transform, while $\mathbf{a}_i = \mathbf{Y}^\dagger \mathbf{W}\boldsymbol{\nu}_i$ is a six-component transform (e.g., for PA4, the $a = 0 - 2$ blocks of $\mathbf{a}_i^a$ would refer to the PA4 f150 band, T, E, and B respectively).

To aid the reader, we have compiled the most important symbols from section 3 and section 4 in appendix B. These symbols are shared among the three noise models; specific differences in their meaning or implementation are explained in the main text.

4.2 Tiled noise model

Previous ACT analyses (e.g. [55]) noted that the stripy noise pattern appeared to be spatially constant across the released maps. In ACT DR4, the relatively small (compared to DR6) sky footprint of each mapped region was referred to as a sky "patch" [2, 58], and thus downstream analyses, notably [19] and [57], built a sparse, 2D Fourier space model of the noise covariance in each patch. In addition, these analyses weighed the Fourier space models by the mapmaker inverse-variance to capture the spatial dependence of the map depth. Thus, within a given patch, the ACT DR4 noise models captured isotropic noise correlations (the "radial" direction in 2D Fourier space), correlated frequencies (by taking cross-spectra in equation (4.1)), map depth inhomogeneity (by inverse-variance weighting difference maps), and spatially-*constant* anisotropy (constant across the patch).

The main goal of the tiled noise model is to reproduce the spatially-*varying* anisotropy in the ACT noise, as shown in figure 4. Our model is based on the "constant correlation" model of [63], which built an inverse-variance weighted, 2D Fourier power spectrum in interleaved $4° \times 4°$ tiles that are cut out of the larger ACT maps. In this sense, the noise model in each tile resembles that for each DR4 patch. We highlight important changes from [63], but for further background, we refer the reader to section 4 of that paper, especially figures 7 and 10–12.

4.2.1 Model estimation

We outline how to measure the tiled noise model, starting from a set of difference maps, $\boldsymbol{\nu}_i$. We follow a similar algorithm as [63], but make a key change in step 2 to improve performance. This also establishes a general prescription for a class of Gaussian map-based noise models, since the wavelet and directional noise models will proceed analogously.

1. As discussed above, we first want to filter the noise maps to maximize their stationarity. But we know, *a priori*, that their covariance is not translationally invariant in map space, because the map depth has spatial structure. Thus, before anything else, we remove the covariance's dependence on absolute sky position — due to the map depth, at least — by multiplying $\boldsymbol{\nu}_i$ with $\sqrt{\mathbf{h}_i}$. We already saw the effect of this in equation (3.10): the per-pixel variance of $\boldsymbol{\nu}_i * \sqrt{\mathbf{h}_i}$ (on small scales) is uniform.[16] We introduce the matrix form of the operation $\boldsymbol{\nu}_i * \sqrt{\mathbf{h}_i}$:

$$\mathbf{H}_i^{\frac{1}{2}} \boldsymbol{\nu}_i \tag{4.2}$$

---

[16] Equations (3.6) and (3.10) are together equivalent to the first step of the noise prescription of [19]: $\mathbf{d}_i * \sqrt{\mathbf{w}_i}$.

where $\mathbf{H}_i^{1/2}$ is a square diagonal matrix whose diagonal entries are populated by $\sqrt{\mathbf{h}_i}$. Thus, $\mathbf{H}_i^{1/2}$ is a "stationarity filter" acting on $\boldsymbol{\nu}_i$.

If we were strictly following the [63] prescription, we would next break up $\mathbf{H}_i^{1/2}\boldsymbol{\nu}_i$ into small tiles and evaluate their 2D Fourier power spectra. Doing so in the context of this paper, however, would introduce some negative side-effects. As we will see in step 3, the tiling of the maps essentially consists of multiplying the maps with small masks. Multiplication by these local masks in map space effects a convolution by a broad kernel in harmonic space. This would result in a mode-coupling bias — power being "smeared" across scales — since the noise power spectra are very steep. Therefore, we first perform a "demodulation" step, in which we divide-out the isotropic part of the noise power spectrum (the curves in the top panel of figure 3) prior to tiling. This results in much flatter power spectra that are less sensitive to mode-coupling. In appendix H, we demonstrate the degree to which mode-coupling induced bias can dominate the tiled noise model, especially for large-scale polarization, without the inclusion of the demodulation step. We proceed as follows.

2. First, we measure the noise power pseudospectra within a large, smooth "pseudospectrum estimate mask", $\boldsymbol{\mu}_{\mathrm{est}}$.[17] That is, we measure all the auto- and cross-pseudospectra of the vector $\boldsymbol{\mu}_{\mathrm{est}} * \sqrt{\mathbf{h}_i} * \boldsymbol{\nu}_i$, which we store in the object $\mathbf{C}_{\mathrm{est}}$. For example, for PA5, $\mathbf{C}_{\mathrm{est}}^{0,3}$ (the $a = 0$, $b = 3$ part of $\mathbf{C}_{\mathrm{est}}$) is the f090 T$\times$ f150 T cross-pseudospectrum: the only off-diagonal parts of $\mathbf{C}_{\mathrm{est}}$ are the frequency and polarization crosses indexed by a and b. Thus, one can also think of $\mathbf{C}_{\mathrm{est}}$ as containing $N_{a_{lm}}$ 6×6 matrices at each harmonic mode. We note that $\mathbf{C}_{\mathrm{est}}$ is a symmetric matrix. Then we construct $\mathbf{C}_{\mathrm{est}}^{-1/2}$, where the matrix exponent acts on each 6×6 matrix of $\mathbf{C}_{\mathrm{est}}$ separately. Finally, we filter the weighted noise maps, $\mathbf{H}_i^{1/2}\boldsymbol{\nu}_i$, by $\mathbf{C}_{\mathrm{est}}^{-1/2}$ in harmonic space:

$$\mathbf{Y}\mathbf{C}_{\mathrm{est}}^{-\frac{1}{2}}\mathbf{Y}^{\dagger}\mathbf{W}\mathbf{H}_i^{\frac{1}{2}}\boldsymbol{\nu}_i. \tag{4.3}$$

This new vector is still in map space, and it still has size $6N_{\mathrm{pix}}$. On small scales, it has uniform per-pixel variance. But now, at least as measured within the apodized mask $\boldsymbol{\mu}_{\mathrm{est}}$, it also has flat autospectra,[18] and has been *decorrelated* — its cross-spectra have been nulled. As discussed, the goal of flattening the spectra is to mitigate mode-coupling when we tile these maps in step 3. The benefit of decorrelating them here will become apparent when we outline the algorithm of drawing a noise simulation: in that case, we will start with a map of easy-to-produce white noise, and by applying the inverse of $\mathbf{C}_{\mathrm{est}}^{-1/2}$ — that is, $\mathbf{C}_{\mathrm{est}}^{1/2}$ — we will emerge with a map with properly-correlated spectral components.

3. After the demodulation step, our goal is still to filter the noise maps to maximize their stationarity. Thus, we break the noise maps into a set of small tiles as motivated by the central assumption of the tiled noise model: that the stripy pattern within small sky patches is constant, but varies over larger distances. We follow the same tiling scheme as [63], wherein the ACT footprint is divided into $M \sim \mathcal{O}(1, 500)$ equal-sized square tiles with a $4°$ pitch. In practice, each tile is actually larger than $4° \times 4°$ to allow for a

---

[17]We need to use $\boldsymbol{\mu}_{\mathrm{est}}$ because the edges of the ACT footprint are not smooth, so the unmasked pseudospectrum would have undesired ringing artifacts. We discuss our choice of mask in 4.5.3.

[18]The power spectrum of the filtered map as measured in small patches of the sky may not be flat, especially in regions outside $\boldsymbol{\mu}_{\mathrm{est}}$, but should have reduced dynamic range compared to figure 3, see appendix H.
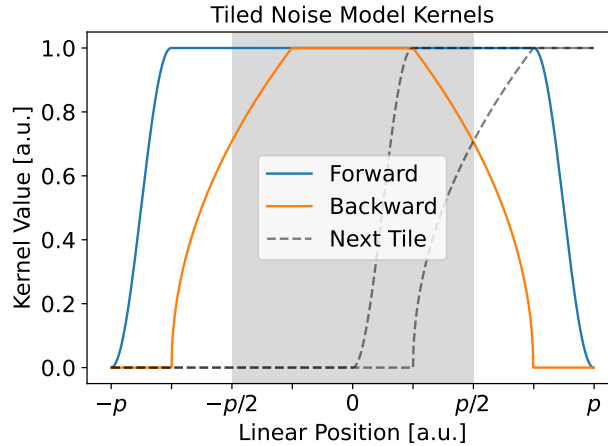
Figure 6. A 1D cross-sectional profile of the "forward" and "backward" tile kernel for a tile of pitch p. The grey band has width p and is centered on the tile. In this paper, we use $p = 4°$. The forward and backward profiles smoothly extend into adjacent tiles; the dashed lines indicate the overlapping kernels for the tile immediately to the right. The full 2D tile kernel is formed by the outer product of perpendicular 1D profiles.

1° cosine apodization around its border.[19] As a result, adjacent tiles partially overlap. We fold the apodization into the definition of the tile "profile", or tile "kernel": cutting a tile out of a map consists of multiplying the map by the tile kernel. A cross-section of an arbitrary tile kernel is shown in figure 6, displaying the apodized edge and overlap with neighboring kernels. We name this operation a "forward tiling transform"; the "backward" transform, which combines tiles back into one map, will be defined when we draw simulations.

We write the matrix form of the forward tiling transform. Each tile has significantly fewer pixels than the full map: $N_{\text{pix,tile}} \ll N_{\text{pix}}$. Thus, the first step in applying a given forward tile — say, tile j — to a map vector is to "crop" the map vector, retaining only those pixels contained in tile j. The next step is to multiply the cropped map vector by the apodized tile kernel itself. Together, these operations look like:

$$\boldsymbol{\tau}_{j,f}\mathbf{P}_j \tag{4.4}$$

where $\mathbf{P}_j$ is a $N_{\text{pix,tile}} \times N_{\text{pix}}$ cropping matrix consisting of $N_{\text{pix,tile}}$ rows of unique unit vectors that select the jth tile pixels, and $\boldsymbol{\tau}_{j,f}$ is a $N_{\text{pix,tile}} \times N_{\text{pix,tile}}$ forward tile matrix, a square diagonal matrix with the jth tile kernel values along its diagonal. The forward tile transform then consists of all of these tile operations acting on the same map vector:

$$\mathcal{T}_f = \begin{pmatrix} \boldsymbol{\tau}_{0,f}\mathbf{P}_0 \\ \boldsymbol{\tau}_{1,f}\mathbf{P}_1 \\ \vdots \\ \boldsymbol{\tau}_{M-1,f}\mathbf{P}_{M-1} \end{pmatrix} \tag{4.5}$$

---

[19]As with $\boldsymbol{\mu}_{\text{est}}$, the apodization prevents strong ringing when taking the per-tile DFT in the next step. Some tiles extend past the edge of the ACT footprint. For such a tile, we simply redefine its new border to be the intersection of the old border and the ACT footprint, and apodize the new border instead.

where $\mathcal{T}_f$ is a block-column-vector of M forward tiles. Putting it all together, our input noise maps have been transformed into weighted, demodulated, tiled noise maps:

$$\mathcal{T}_f \mathbf{Y} \mathbf{C}_{\mathrm{est}}^{-\frac{1}{2}} \mathbf{Y}^\dagger \mathbf{W} \mathbf{H}_i^{\frac{1}{2}} \boldsymbol{\nu}_i. \tag{4.6}$$

4. Assuming that within each tile the noise covariance is stationary, we take our cue from the cross-correlation theorem (equation (4.1)). That is, by transforming to Fourier space, we will assume that each Fourier mode in the resulting vector is *uncorrelated* with all other Fourier modes:

$$\mathbf{u}_i \equiv \mathbf{F} \mathcal{T}_f \mathbf{Y} \mathbf{C}_{\mathrm{est}}^{-\frac{1}{2}} \mathbf{Y}^\dagger \mathbf{W} \mathbf{H}_i^{\frac{1}{2}} \boldsymbol{\nu}_i \tag{4.7}$$

where $\mathbf{u}_i$ is the weighted, demodulated, tiled, 2D Fourier projection of the map-based noise, and $\mathbf{F}$ is understood to be a square block-diagonal matrix of DFTs (one for each tile). For each Fourier mode, however, we still retain the six correlated map fields: two frequency bands and three Stokes components.

5. With the construction of $\mathbf{u}_i$, most of the hard work is complete. The last steps use $\mathbf{u}_i$ to build a sparse (diagonal) covariance matrix, just as in equation (4.1). For each tile, we evaluate its 2D Fourier noise power spectra:

$$\mathcal{N}(\mathbf{k})_{i,j}^{ab} \delta(\mathbf{k} - \mathbf{k}') = \frac{1}{f_{j,2}} \Re\left(\mathbf{u}_{i,j}^a(\mathbf{k}) \mathbf{u}_{i,j}^{b*}(\mathbf{k}) \right) \tag{4.8}$$

where $\mathbf{k}$ is a given 2D Fourier mode, $\Re$ denotes the real part (relevant when $a \neq b$), $\mathbf{u}_{i,j}^a(\mathbf{k})$ is the jth tile and a^{th} component of our transformed noise map for split i, and $f_{j,2}$ corrects for the loss of power from the apodization in the tile [19, 57, 63]:

$$f_{j,2} = \frac{1}{N_{\mathrm{pix,tile}}} \mathrm{Tr}[\boldsymbol{\tau}_{j,f}^\dagger \boldsymbol{\tau}_{j,f}] \tag{4.9}$$

which is just the average value of the squared tile kernel. Thus, in each tile, the covariance matrix $\mathcal{N}_{i,j}$ is block-diagonal in each Fourier mode $\mathbf{k}$, with a single block being a 6×6 matrix $\mathcal{N}(\mathbf{k})_{i,j}$. The total covariance matrix $\mathcal{N}_i$ is block-diagonal in tiles, with a single block being $\mathcal{N}_{i,j}$:

$$\mathcal{N}_i = \begin{pmatrix} \mathcal{N}_{i,0} & 0 & \dots & 0 \\ 0 & \mathcal{N}_{i,1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathcal{N}_{i,M-1} \end{pmatrix}. \tag{4.10}$$

This highly diagonal form is the sparse representation we originally sought: it has $\mathcal{O}(N_{\mathrm{pix}})$ nonzero elements near its diagonal, rather than being dense with $\mathcal{O}(N_{\mathrm{pix}}^2)$ nonzero elements.

6. As in [19, 57, 63], we smooth each tile's power spectrum estimate $N(\mathbf{k})_{i,j}^{ab}$ over nearby modes $\mathbf{k}$ to reduce sample variance. Unlike the smoothing procedure for the two wavelet models (discussed in section 4.5.4 and appendix G), for the tiled model we cannot reliably distinguish between random fluctuations and true features in the tiled 2D power spectra. We thus select the same smoothing width in Fourier space as was used
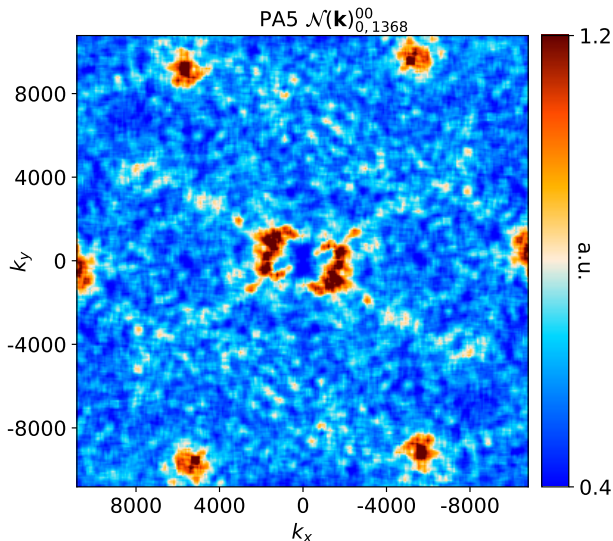
Figure 7. The tiled 2D Fourier power spectrum for PA5 f090, first split ($i = 0$). This particular power spectrum corresponds to tile $j = 1,368$ in a model bandlimited to $\ell_{\max} = 10,800$, and we see all Fourier modes $\mathbf{k}$ for the TT ($a = b = 0$) component. As in figure 4, we can visualize the local anisotropic structure in this tile, including an "x-like" scan strategy and detector-detector correlations at the image perimeter (small scales). More examples are visible in [63], figure 11.

in [57] and [63], $\Delta k = 400$ in both the x- and y-directions. We perform the smoothing via convolution with a tophat kernel directly applied to the 2D power spectra.[20] The smoothing width is fixed both within a given tile, and for each tile. An example smoothed, tiled 2D Fourier power spectrum is shown in figure 7.

7. Lastly, in preparation for drawing noise simulations, we note that the actual objects we save to disk are not the tiled 2D Fourier covariance matrices, but rather their matrix square-root. This is the matrix analog of obtaining the standard deviation from the variance in preparation for sampling a scalar Gaussian random variable. Thus, the object we save to disk is actually $\mathcal{N}_i^{1/2}$. As in step 2, the only off-diagonal parts of $\mathcal{N}_i$ are the frequency and polarization crosses indexed by a and b. Therefore, the matrix exponent acts on each 6×6 matrix $\mathcal{N}(\mathbf{k})_{i,j}$ separately. Some examples of $\mathcal{N}(\mathbf{k})_{i,j}^{1/2}$ are visible in figure 8. Because it is also used in drawing simulations, we save $\mathbf{C}_{\text{est}}^{1/2}$ to disk as well.

These steps generate the tiled model square-root covariance $\mathcal{N}_i^{1/2}$ for the purposes of this paper. A schematic depiction of all these steps is given in the top row of figure 8.

4.2.2 Simulations and covariance matrix

The ultimate aim of this paper is to create realistic simulations of the map-based noise. Here, we explain how we draw such a simulation from the model we have just described. Along the way, we also write down the full map-based noise covariance matrix. The following steps are essentially applicable to the wavelet models as well; in section 4.3.2 and section 4.4.2

---

[20]Recall we have already divided-out the steep isotropic part of the power spectrum, so we are smoothing over the *anisotropic* features in $N(\mathbf{k})_{i,j}^{ab}$ *relative* to the isotropic power spectrum.
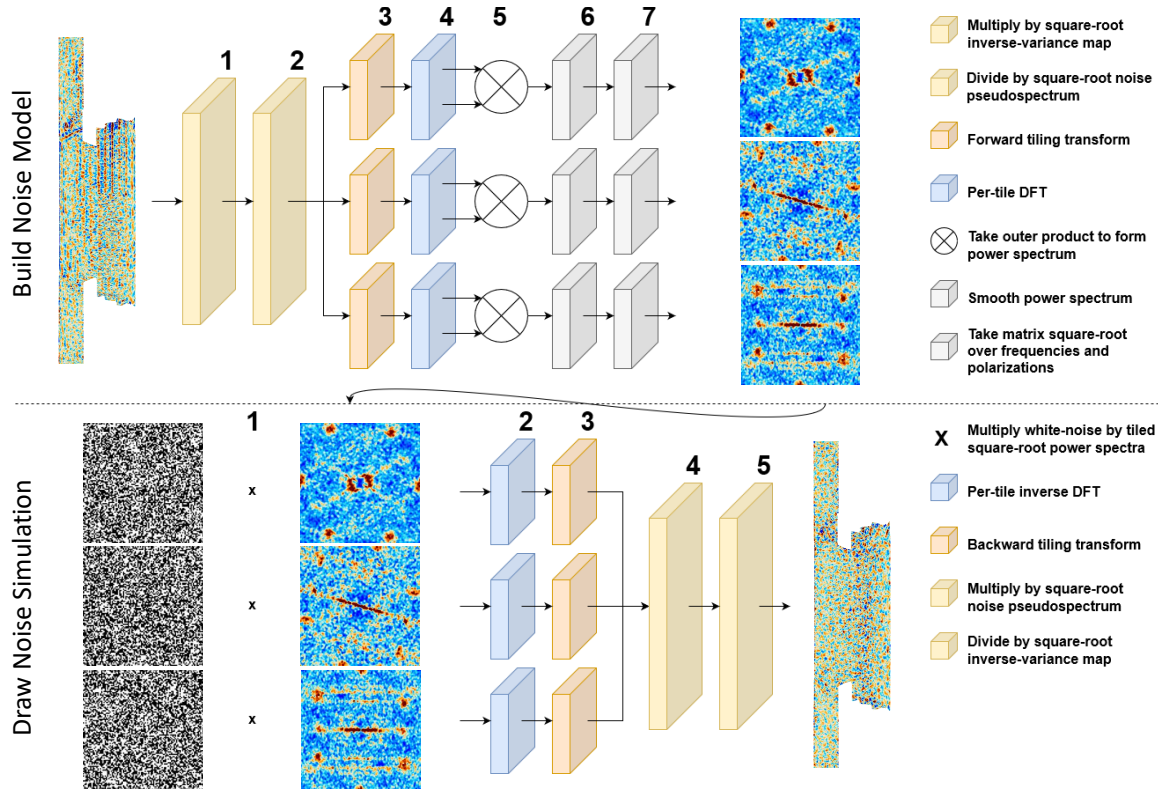
Figure 8. *Top*: the sequence of filters and other operations involved in measuring the sparse covariance in the tiled noise model. Large boxes denote filters applied to the entire map, while small boxes denote filters or operations in the tiled basis. The orange filters represent the "forward" tiling kernels in figure 6. Each operation is labelled by the corresponding step number in the model estimation algorithm. The input difference map is PA5 f090, T ($a = 0$), first split ($i = 0$), and the outputs are 2D square-root Fourier power spectra, TT ($a = b = 0$). We see the noise anisotropy varies by tile. *Bottom*: the sequence of filters and other operations involved in drawing a simulation of the map-based noise in the tiled model. The first column denotes independent, Gaussian white noise. Importantly, the orange filters now represent the "backward" tiling kernels in figure 6, while all other operations revert those of the top row. Each operation is labelled by the corresponding step number in the simulation algorithm. The output is a tiled noise simulation of PA5 f090, T, first split.

we only discuss minor differences from what follows here. Batches of these map-based noise simulations are the main product of this paper.

There are two key steps in extracting a realization of the map-based noise from the tiled model. We start by drawing a realization in tiled 2D Fourier space. Doing so is tractable thanks to the highly diagonal structure of the tiled 2D Fourier covariance, $\mathcal{N}_i$. Then we need to transform the realization back to map space, which essentially means reverting the various operations that built $\mathbf{u}_i$ from $\boldsymbol{\nu}_i$ in equation (4.7).

1. For each tile, draw Gaussian white noise in tiled 2D Fourier space and take the inner product with the square-root covariance:

$$\mathcal{N}_i^{\frac{1}{2}} \boldsymbol{\eta}_i \tag{4.11}$$

where $\boldsymbol{\eta}_i$ is complex-valued white noise with the same shape as $\mathbf{u}_i$. Again, the only nontrivial (off-diagonal) part of this product is in the frequency/polarization index. By

definition, $\langle \boldsymbol{\eta}_i \boldsymbol{\eta}_i^\dagger \rangle = \mathbf{I}$, where $\mathbf{I}$ is the identity. Therefore, $\mathcal{N}_i^{1/2} \boldsymbol{\eta}_i$ is a sample from the tiled covariance:

$$\langle \mathcal{N}_i^{\frac{1}{2}} \boldsymbol{\eta}_i (\mathcal{N}_i^{\frac{1}{2}} \boldsymbol{\eta}_i)^\dagger \rangle = \mathcal{N}_i^{\frac{1}{2}} \langle \boldsymbol{\eta}_i \boldsymbol{\eta}_i^\dagger \rangle \mathcal{N}_i^{\frac{1}{2},\dagger} = \mathcal{N}_i^{\frac{1}{2}} \mathcal{N}_i^{\frac{1}{2},\dagger} = \mathcal{N}_i \tag{4.12}$$

by construction. This demonstrates why we save to disk the square-root covariance.

2. With our tiled 2D Fourier space noise realization in hand, we proceed to revert equation (4.7). So, first we transform out of tiled Fourier space and back to tiled map space with a per-tile inverse DFT:

$$\mathbf{F}^\dagger \mathcal{N}_i^{\frac{1}{2}} \boldsymbol{\eta}_i. \tag{4.13}$$

3. Next, we need to stitch tiles back together into a single map. That is, we need to apply a "backward" tiling transform. At this point in the algorithm, each tile contains an independent realization of its own stripy noise pattern. Therefore, as noted in [63], naively pasting each tile next to one another with a step-function edge would produce sharp discontinuities in the synthesized map. Just like the forward transform, the solution is to apodize and overlap their borders such that the transition across tiles is continuous. A good illustration of this is shown in [63], figure 13. Unfortunately, we cannot just reuse the tile kernels from the forward transform: we should specifically design the apodization to preserve the per-pixel noise variance in the tile overlap. We show the exact constraint that must be satisfied in equation (4.16) and derive it in appendix D, but, in short, the forward tile profile does not satisfy it. Instead, we use the backward profile shown in figure 6, which is the square-root of the bilinear interpolation from [63].

The matrix form of the backward tiling transform is identical to that of the forward tiling transform (equation (4.5)):

$$\mathcal{T}_b = \begin{pmatrix} \boldsymbol{\tau}_{0,b} \mathbf{P}_0 \\ \boldsymbol{\tau}_{1,b} \mathbf{P}_1 \\ \vdots \\ \boldsymbol{\tau}_{M-1,b} \mathbf{P}_{M-1} \end{pmatrix}. \tag{4.14}$$

The only difference is that the diagonal of each $N_{\text{pix,tile}} \times N_{\text{pix,tile}}$ backward tile matrix, $\boldsymbol{\tau}_{j,b}$, contains the values of the backward tile kernels. The projection matrices $\mathbf{P}_j$ are unchanged: by applying $\mathcal{T}_b^\dagger$, they perform a "paste" operation that is the exact opposite of the "crop" operation from the forward transform. Thus we have:

$$\mathcal{T}_b^\dagger \mathbf{F}^\dagger \mathcal{N}_i^{\frac{1}{2}} \boldsymbol{\eta}_i \tag{4.15}$$

where the vector now has the same shape as an input noise map, $\boldsymbol{\nu}_i$, and likewise lives in the original, untiled map space.

As discussed, when stitched together in the synthesized map, the backward tile profiles should preserve the total per-pixel noise variance. For example, for a given pixel belonging only to one tile, it should be the case that the value of that tile's backward kernel is 1 — or else we would artificially increase or decrease that pixel's variance. In general —

including for pixels in an overlapping region between multiple tiles — the *quadrature sum* of all contributing tile kernels should likewise be 1:

$$\mathbf{I} = \sum_{j=0}^{M-1} \mathbf{P}_j^\dagger (\boldsymbol{\tau}_{j,b}^\dagger \boldsymbol{\tau}_{j,b}) \mathbf{P}_j \tag{4.16}$$

where the term in parentheses is just the squared value of the backward tile kernels, and the sum over projection matrices performs the tile "stitching". We derive equation (4.16) more rigorously in appendix D. Continuing with the simulation algorithm:

4. We need to apply the inverse of the demodulation filter from the model estimation. That is, we multiply the simulation in harmonic space by $\mathbf{C}_{\text{est}}^{1/2}$:

$$\mathbf{Y}\mathbf{C}_{\text{est}}^{\frac{1}{2}}\mathbf{Y}^\dagger\mathbf{W}\mathcal{T}_b^\dagger\mathbf{F}^\dagger\mathcal{N}_i^{\frac{1}{2}}\boldsymbol{\eta}_i. \tag{4.17}$$

This reapplies the steep noise auto- and cross-spectra to the simulation.

5. The last step in drawing a noise simulation is to divide by the square-root of the inverse-variance map, or multiply element-wise by $1/\sqrt{\mathbf{h}_i}$. This is analogous to multiplying by the per-pixel standard deviation. In matrix form, a simulation from the tiled noise model is therefore defined as:

$$\boldsymbol{\xi}_i \equiv \mathbf{H}_i^{-\frac{1}{2}}\mathbf{Y}\mathbf{C}_{\text{est}}^{\frac{1}{2}}\mathbf{Y}^\dagger\mathbf{W}\mathcal{T}_b^\dagger\mathbf{F}^\dagger\mathcal{N}_i^{\frac{1}{2}}\boldsymbol{\eta}_i. \tag{4.18}$$

These steps generate a tiled model noise simulation, $\boldsymbol{\xi}_i$, for the purposes of this paper. A schematic depiction of all these steps is given in the bottom row of figure 8. We also have the necessary ingredients to write the full map-based noise covariance matrix in the tiled model. By definition, the map-based noise covariance matrix is:

$$\mathbf{N}_i = \langle \boldsymbol{\xi}_i \boldsymbol{\xi}_i^\dagger \rangle. \tag{4.19}$$

Combining equations (4.12), (4.18), and (4.19) we have:

$$\mathbf{N}_i = \mathbf{H}_i^{-\frac{1}{2}}\mathbf{Y}\mathbf{C}_{\text{est}}^{\frac{1}{2}}\mathbf{Y}^\dagger\mathbf{W}\mathcal{T}_b^\dagger\mathbf{F}^\dagger\mathcal{N}_i\mathbf{F}\mathcal{T}_b\mathbf{W}\mathbf{Y}\mathbf{C}_{\text{est}}^{\frac{1}{2}}\mathbf{Y}^\dagger\mathbf{H}_i^{-\frac{1}{2}} \tag{4.20}$$

where we have used that $\mathbf{H}_i^{-1/2}$, $\mathbf{C}_{\text{est}}^{1/2}$, and $\mathbf{W}$ are Hermitian.

This section has given a detailed accounting of how we both measure the tiled noise model and draw a tiled noise simulation. While thorough, it is an investment that will pay off in our discussion of the two wavelet models: in spite of the wavelets' orthogonal approach to modeling the map-based noise, their actual mechanics are highly similar to the tiled model.

4.3 Isotropic wavelet noise model

Our isotropic wavelet, or "wavelet", noise model especially targets the scale-dependent map depth of the ACT DR6 noise, as shown in figure 5. The model works by taking figure 5 at face value: maps of the per-pixel noise variance appear different depending on the filtered range of angular scales. That is, each range of angular scales could be assigned its own "inverse-variance" map, where each map still operates in the white noise, uncorrelated-pixel limit. This idea is especially well motivated by the observation that the transition from

the small-scale to the large-scale map depth in figure 5 is reasonably smooth (recall that the bottom panel shows the morphology after normalizing by the mapmaker inverse-variance). In other words, we do not need especially high-resolution "filters" in harmonic space to capture the simultaneous scale- and spatial-dependence of the map depth. We therefore assume that the set of scale-dependent inverse-variance maps forms a complete and sparse description of the map-based noise.

Applying the core principle of the cross-correlation theorem (equation (4.1)), a pixel-wise uncorrelated covariance in map space inspires us to seek stationary regions in harmonic space. We isolate these regions by tiling the harmonic plane, where now we call the harmonic space kernels "wavelets": a set of smooth, interleaved profiles defined over ℓ, each of which filters for a compact range of angular scales.[21] Thus, the swapping of map space and harmonic space is the main *algorithmic* difference between the tiled model and the wavelet model, and we can proceed nearly identically as in section 4.2. We first prescribe how we construct a sparse representation of the map-based noise covariance under the wavelet model in section 4.3.1. For reference, this procedure is similar to the construction of the total data covariance in needlet ILC (NILC) analyses (e.g. [23, 27, 70]). Then, in section 4.3.2, we outline how we draw a simulation from the wavelet model.

4.3.1 Model estimation

As before, the goal of the model estimation algorithm is to first filter the ACT DR6 noise maps to promote their stationarity, and then transform them to permit a sparse representation of the covariance matrix. In the wavelet model, we work in harmonic space first and then transform into map space.

1. We start by considering the harmonic transform of the noise maps. Thus, our input noise vectors are given by:

$$\mathbf{Y}^{\dagger}\mathbf{W}\boldsymbol{\nu}_i \tag{4.21}$$

   which is now a $6N_{a_{lm}}$ element vector instead of a $6N_{\mathrm{pix}}$ element vector.

2. Just as in the tiled model, at this stage, we have *a priori* knowledge that the noise covariance is not translationally invariant in harmonic space. Instead, the noise variance depends on the absolute position in harmonic space — that is, depends on ℓ — due to the steep noise power spectra (figure 3). Therefore, we still need to apply a "stationarity filter", only this time in harmonic space instead of map space. We perform the same filtering operation as in section 4.2.1, step 2, wherein we divide-out the square-root noise spectra[22] as measured in the same pseudospectrum estimate mask, $\boldsymbol{\mu}_{\mathrm{est}}$:

$$\mathbf{C}_{\mathrm{est}}^{-\frac{1}{2}}\mathbf{Y}^{\dagger}\mathbf{W}\boldsymbol{\nu}_i. \tag{4.22}$$

   Whereas in the tiled model, this operation functioned as a demodulation filter, now the flattened power spectra help promote the noise stationarity in harmonic space. Again, we further explore the necessity of this filter in appendix H.

---

[21]Technically speaking, there is not an exact analog of equation (4.1) in harmonic space; instead, we use it as an approximately true "guiding principle".

[22]The $\mathbf{C}_{\mathrm{est}}$ measured in the wavelet model will not be the same as the tiled model, since in the tiled model we had already multiplied the noise maps by the square-root inverse-variance before measuring the spectra.

In principle, we could now perform the harmonic space equivalent of a "demodulation" step to avoid mode-coupling. Recall, in the (map space) tiled model, this involved a filter operating in harmonic space — dividing-out the square-root noise pseudospectra. The effect of this was to suppress mixing power across angular scales when tiling the maps. Therefore, for the (harmonic space) wavelet model, demodulation would involve a filter operating in *map space* — namely, multiplying by the square-root inverse-variance maps (as in equation (3.10)). Analogous to the tiled model, this would suppress mixing power across regions of the map when breaking the harmonic plane into wavelets in the next step: multiplication by a local wavelet profile in harmonic space effects a convolution by a broad kernel in map space. Nevertheless, in practice we found this demodulation step was not necessary for performance of the wavelet model, and we chose to forego it here. Although we do not include it, we discuss the drawbacks and benefits of this demodulation step further in section 6.

3. After flattening the spectra, our goal is still to isolate stationary regions in the transformed noise. As discussed, we achieve this by breaking the noise into a set of "wavelets", which are analogous to tiles defined in harmonic space. For this purpose we use the "scale-discrete" wavelets of [52, 85] with a scaling parameter $\lambda = 1.3$ (λ sets the logarithmic wavelet width), lowest wavelet scale $J_0 = 9$ (which sets the $\ell_{\max}$ of the lowest-ℓ wavelet), and highest wavelet scale $J = 33$ (which determines the overall number of wavelets). These parameters generate the $M = 26$ wavelet kernels shown in the top panel of figure 9.[23] The wavelet log-spacing is well suited to the roughly power-law scaling of the noise spectra. In addition, the smooth overlap of the wavelet profiles plays a similar role as the apodized tile edges in section 4.2.1, step 3: it prevents strong ringing of the *map-space* wavelet kernels when taking the per-wavelet SHT in the next step. Said another way, it sacrifices some of each wavelet's resolution in harmonic space to increase its resolution in map space. The bottom panels of figure 9 demonstrate just that: the map-space wavelet kernels are compact, and can therefore resolve local noise features in map space. We also observe that because the higher-ℓ wavelets are wider in harmonic space, their map-space kernels are commensurately tighter. Thus, the wavelet decomposition is capable of simultaneously localizing noise properties over angular scale and sky position.

We again denote this operation a "forward tiling transform". As before, the first step in applying a given wavelet profile to a map vector is to "crop" the noise vector. Only now, we retain all harmonic modes up to the wavelet $\ell_{\max}$, which is to say we *bandlimit* the noise vector for each wavelet. Then, we multiply the set of bandlimited vectors by the corresponding wavelet profile. For a given wavelet j, the matrix form is identical to equation (4.4), where $\mathbf{P}_j$ is a $N_{a_{lm},\mathrm{wav}} \times N_{a_{lm}}$ bandlimiting matrix, and $\boldsymbol{\tau}_{j,f}$ is a $N_{a_{lm},\mathrm{wav}} \times N_{a_{lm},\mathrm{wav}}$ square diagonal matrix with the jth wavelet kernel values along its diagonal. Then the forward tiling transform, $\mathcal{T}_f$ has the exact same form as equation (4.5). Our input noise maps thus become transformed, weighted, tiled noise maps:

$$\mathcal{T}_f \mathbf{C}_{\mathrm{est}}^{-\frac{1}{2}} \mathbf{Y}^\dagger \mathbf{W} \boldsymbol{\nu}_i. \tag{4.23}$$

4. We again are guided by the cross-correlation theorem (equation (4.1)): we now transform into map space, where we assume that each pixel in the resulting vector is *uncor-*

---

[23]We augment the highest-ℓ wavelet kernel to extend to arbitrarily high-ℓ at a value of 1.
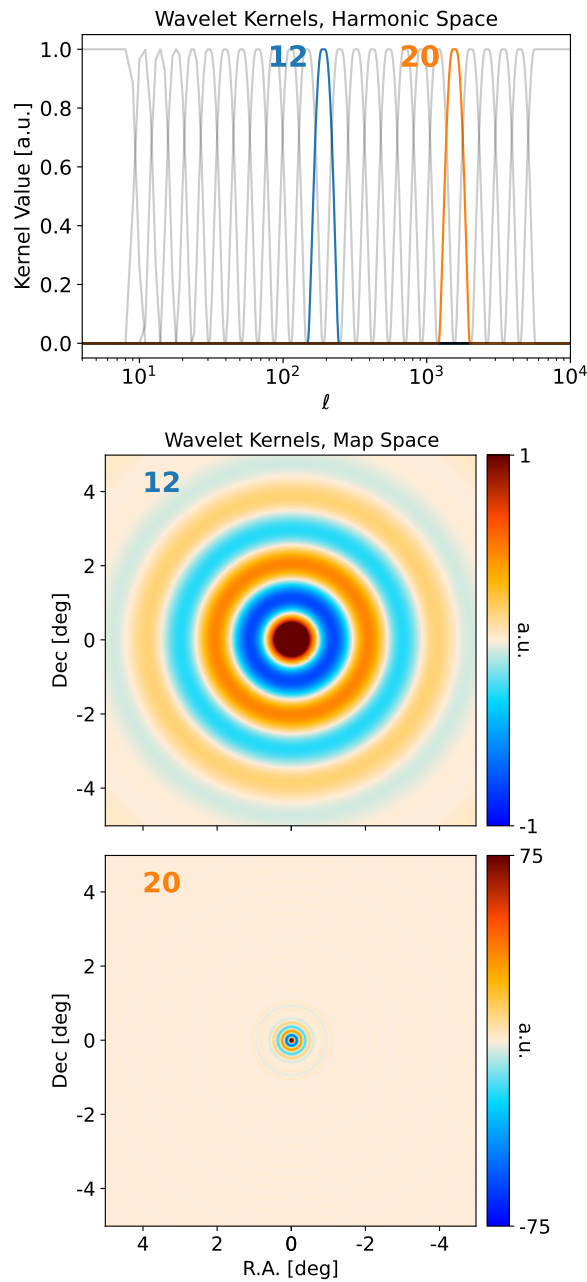
Figure 9. *Top*: scale-discrete wavelet kernels [52, 85] defined in harmonic space. Kernels with indices $j = 12$ and $j = 20$ are highlighted. *Middle*: kernel $j = 12$ in map space, with support in ℓ of $\sim 100{-}200$. As expected, this kernel captures scales on the order of $\sim 1°{-}2°$. *Bottom*: kernel $j = 20$ in map space, with support in ℓ of $\sim 1,000{-}2,000$. As expected, this kernel captures scales on the order of $\sim 0.1°{-}0.2°$. Its wider width in harmonic space translates to a narrower width in map space.

related with all other pixels:

$$\mathbf{u}_i \equiv \mathbf{Y}\mathcal{T}_f \mathbf{C}_{\mathrm{est}}^{-\frac{1}{2}} \mathbf{Y}^\dagger \mathbf{W}\boldsymbol{\nu}_i. \tag{4.24}$$

Thus, $\mathbf{u}_i$ contains a set of M bandlimited "wavelet maps", where each map has six

correlated frequency/polarization components, but whose pixels are uncorrelated.

5. Using $\mathbf{u}_i$, we evaluate the pixel-diagonal wavelet power maps:

$$\mathcal{N}(\mathbf{x})_{i,j}^{ab}\delta(\mathbf{x}-\mathbf{x}') = \frac{1}{f_{j,2}(\mathbf{x})}\left(\mathbf{u}_{i,j}^a(\mathbf{x})\mathbf{u}_{i,j}^{b*}(\mathbf{x})\right) \qquad (4.25)$$

where $\mathbf{x}$ is a given wavelet map pixel, $\mathbf{u}_{i,j}^a(\mathbf{x})$ is the jth wavelet map and a^{th} component of our transformed noise vector evaluated at $\mathbf{x}$, and $f_{j,2}(\mathbf{x})$ corrects for the power suppression due to the modes lost when multiplying by each wavelet profile [25]:

$$f_{j,2}(\mathbf{x}) = \frac{A(\mathbf{x})}{4\pi}\text{Tr}[\boldsymbol{\tau}_{j,f}^\dagger\boldsymbol{\tau}_{j,f}] \qquad (4.26)$$

where A, the pixel area in steradians, accounts for the spherical geometry of the wavelets. The total wavelet covariance matrix, $\mathcal{N}_i$, is thus the same form as equation (4.10). Each block along the diagonal, $\mathcal{N}_{i,j}$, is itself block-diagonal in map pixel $\mathbf{x}$, with a single block being a 6×6 matrix $\mathcal{N}(\mathbf{x})_{i,j}$; the elements of $\mathcal{N}(\mathbf{x})_{i,j}$ are given by equation (4.25). Again, this highly diagonal form is our desired sparse representation, containing $\mathcal{O}(N_{a_{lm}})$ nonzero elements near its diagonal.
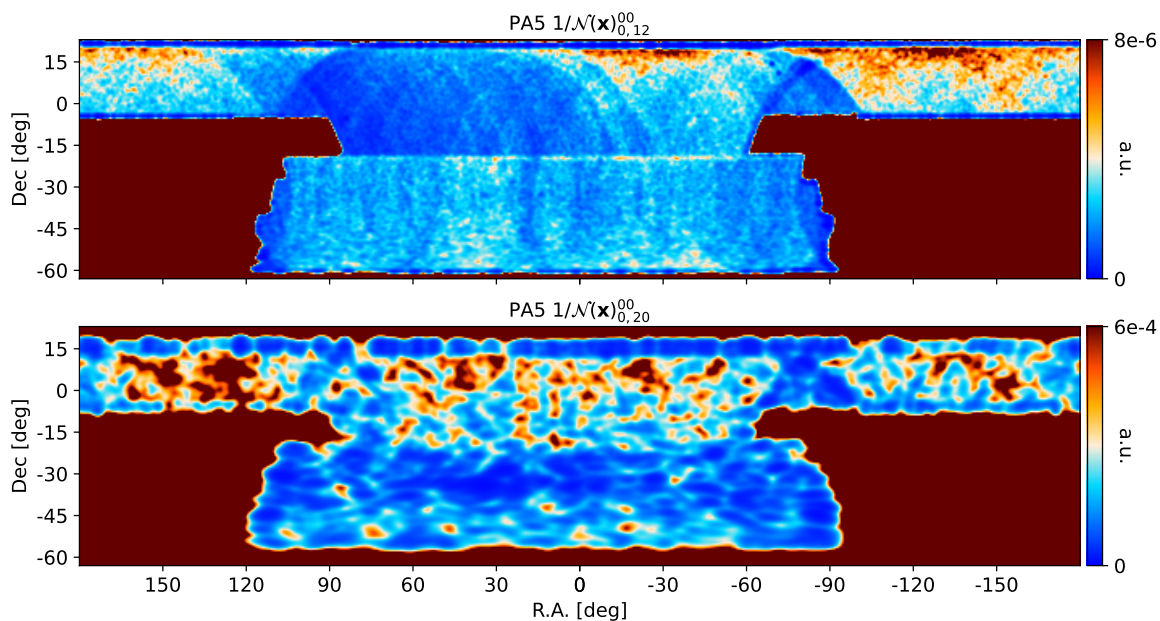
Figure 10. *Top (Bottom)*: the inverse of a wavelet noise power map corresponding to the high-ℓ, $j = 20$ (low-ℓ, $j = 12$) wavelet kernels highlighted in figure 9. The particular component is PA5 f090, TT ($a = b = 0$), first split ($i = 0$). As expected, when compared to figure 5, we see the small-scale (large-scale) wavelet inverse-variance is suggestive of the morphology of the mapmaker inverse-variance (cross-linking) map.

6. We smooth each wavelet power map $\mathcal{N}(\mathbf{x})_{i,j}^{ab}$ over nearby pixels $\mathbf{x}$ to reduce sample variance. Some examples of smoothed wavelet power maps are shown in figure 10. We see the wavelet model is suggestive of cross-linking-like noise morphology at large scales, and inverse-variance-like noise morphology at small scales.

   Unlike the tiled model, we do not use the same smoothing width for all wavelets. Instead, we select a Gaussian smoothing kernel with a FWHM of $f(\ell_{\max})\pi/\ell_{\max}$ radians, where f is some function of ℓ and $\ell_{\max}$ is the maximum support of the given wavelet. Because the $\ell_{\max}$ of our SHTs are set to the Nyquist bandlimit of the corresponding map pixelization, $\pi/\ell_{\max}$ for a given wavelet is the pixel size (in radians) of the corresponding wavelet map. Thus, f counts the number of pixels in the wavelet power maps across the smoothing FWHM. As shown in section 4.5.4, figure 16, f takes on values of $\mathcal{O}(10)$. The choice of $f(\ell)$ is discussed further in section 4.5.4, as well as in appendix G.

7. Lastly, as before, we save to disk the matrix square-root of $\mathcal{N}_i$ — that is, $\mathcal{N}_i^{1/2}$ — and $\mathbf{C}_{\mathrm{est}}^{1/2}$. Some examples of $\mathcal{N}(\mathbf{x})_{i,j}^{1/2}$ are visible in figure 11.

These steps generate the wavelet model square-root covariance $\mathcal{N}_i^{1/2}$ for the purposes of this paper. A schematic depiction of all these steps is given in the top panel of figure 11.

4.3.2 Simulations and covariance matrix

Just as the wavelet noise model estimation algorithm followed directly from section 4.2.1, so too does its simulation algorithm carry over from section 4.2.2. That is, the procedure for
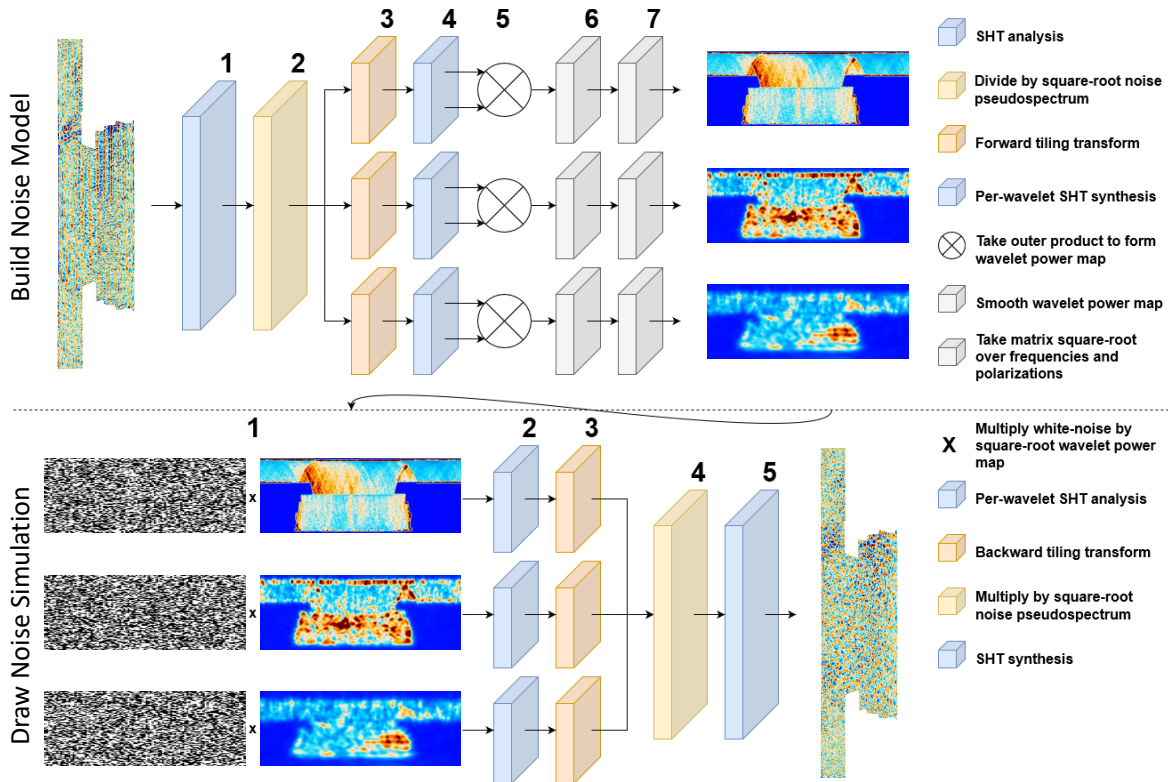
Figure 11. *Top*: the sequence of filters and other operations involved in measuring the sparse covariance in the wavelet noise model. Large boxes denote filters or operations applied to the entire map, while small boxes denote filters or operations after tiling the harmonic plane. The orange filters represent the "forward" wavelet kernels in figure 9 ("forward" and "backward" kernels are the same in this model). Each operation is labelled by the corresponding step number in the model estimation algorithm. The input difference map is PA5 f090, T ($a = 0$), first split ($i = 0$), and the outputs are square-root wavelet power maps, TT ($a = b = 0$). We see the spatial variation in map depth also varies by wavelet scale, with high-ℓ to low-ℓ running top to bottom. *Bottom*: the sequence of filters and other operations involved in drawing a simulation of the map-based noise in the wavelet model. The first column denotes independent, Gaussian white noise. The orange filters represent the "backward" wavelet kernels in figure 9, while all other operations revert those of the top row. The output is a wavelet noise simulation of PA5 f090, T, first split.

generating a noise simulation from the wavelet model consists of first drawing a realization in wavelet map space, and then reverting the operations that transformed $\boldsymbol{\nu}_i$ into $\mathbf{u}_i$ in equation (4.24).

1. For each wavelet, draw Gaussian white noise in wavelet map space and take the inner product with the square-root covariance:

$$\mathcal{N}_i^{\frac{1}{2}} \boldsymbol{\eta}_i \tag{4.27}$$

where $\boldsymbol{\eta}_i$ is real-valued white noise with the same shape as $\mathbf{u}_i$.

2. Next, transform out of wavelet map space and back to wavelet harmonic space with a

per-wavelet SHT:

$$\mathbf{Y}^{\dagger}\mathbf{W}\mathcal{N}_{i}^{\frac{1}{2}}\boldsymbol{\eta}_{i}. \tag{4.28}$$

That is, this vector contains the SHT of the $M = 26$ independent wavelet noise realizations.

3. Just as in a tiled noise simulation, we now need to stitch the wavelet realizations back together, only this time the stitching occurs in harmonic space. The same considerations carry over from the tiled model as well: we should avoid naively pasting each wavelet next to one another, which would result in sharp discontinuities in the synthesized harmonic transform. Thus, we first multiply the wavelets by smoothly tapered profiles in harmonic space. For this purpose, we reuse the same profiles, or "kernels", from the forward transform, as shown in figure 9. As before, we also require that the sum over wavelets of the per-harmonic-mode noise variance is preserved; equivalently, that the wavelet kernels satisfy equation (4.16). Conveniently, the wavelets of [52, 85] do so by construction. We then simply add up the 26 smoothly tapered wavelet realizations in harmonic space, resulting in a single harmonic transform. Together, this constitutes the "backward tiling transform".

The matrix form of the backward tiling transform is identical to equation (4.14). Since we have reused the same wavelet profiles as in the forward transform ($\mathcal{T}_{b} = \mathcal{T}_{f}$), we therefore have:

$$\mathcal{T}_{f}^{\dagger}\mathbf{Y}^{\dagger}\mathbf{W}\mathcal{N}_{i}^{\frac{1}{2}}\boldsymbol{\eta}_{i} \tag{4.29}$$

where the vector now has the same shape as the harmonic transform of $\boldsymbol{\nu}_{i}$, and likewise lives in the original, "untiled" harmonic space.

4. We apply the inverse of the stationarity filter from the model estimation. That is, we reapply the square-root noise auto- and cross-spectra directly in harmonic space:

$$\mathbf{C}_{\mathrm{est}}^{\frac{1}{2}}\mathcal{T}_{f}^{\dagger}\mathbf{Y}^{\dagger}\mathbf{W}\mathcal{N}_{i}^{\frac{1}{2}}\boldsymbol{\eta}_{i}. \tag{4.30}$$

5. Finally, transforming the entire simulation back into map space completes the noise simulation:

$$\boldsymbol{\xi}_{i} = \mathbf{Y}\mathbf{C}_{\mathrm{est}}^{\frac{1}{2}}\mathcal{T}_{f}^{\dagger}\mathbf{Y}^{\dagger}\mathbf{W}\mathcal{N}_{i}^{\frac{1}{2}}\boldsymbol{\eta}_{i}. \tag{4.31}$$

These steps generate a wavelet model noise simulation, $\boldsymbol{\xi}_{i}$, for the purposes of this paper. A schematic depiction of these steps is given in the bottom panel of figure 11. As before, it is easy to use equation (4.31) to write the full map-based noise covariance matrix in the wavelet model:

$$\mathbf{N}_{i} = \mathbf{Y}\mathbf{C}_{\mathrm{est}}^{\frac{1}{2}}\mathcal{T}_{f}^{\dagger}\mathbf{Y}^{\dagger}\mathbf{W}\mathcal{N}_{i}\mathbf{W}\mathbf{Y}\mathcal{T}_{f}\mathbf{C}_{\mathrm{est}}^{\frac{1}{2}}\mathbf{Y}^{\dagger} \tag{4.32}$$

where we have used that $\mathbf{C}_{\mathrm{est}}^{1/2}$ and $\mathbf{W}$ are Hermitian.

This completes our discussion of the isotropic wavelet model. The direct parallels to the tiled model construction are apparent. By tiling harmonic space instead of map space, the wavelet model is well-adapted to capture the scale-dependence of the map depth; however, by construction, the isotropic wavelets are circular in map space (figure 9) — they are insensitive to any local noise stripiness.

4.4 Directional wavelet noise model

Our directional wavelet, or "directional", noise model in principle captures all enumerated aspects of the ACT map-based noise, including the spatially-varying stripy patterns and scale-dependent map depth, simultaneously. Its key feature is a decomposition using a set of directional wavelets: each wavelet filters for a particular range of angular scales (like the "isotropic" wavelet model), as well as a particular direction on the sky (unlike the "isotropic" wavelet model). That is, by assuming the noise is stationary within a "directional wavelet" with both a well-defined angular scale *and* direction, the decomposition naturally leads to scale- and direction-dependent "inverse-variance" maps. Moreover, because the main difference between the wavelet model and directional model is just the definition of their wavelet profiles, their model estimation and simulation algorithms are nearly identical. The directional model can therefore be viewed as a generalization of the wavelet model.

The directional model's only algorithmic change is that we build the wavelets in 2D Fourier space instead of harmonic space.[24] Doing so admits speed improvements, increased flexibility in wavelet design, and generally simpler implementation. One consequence of defining wavelets in Fourier space is that their map space kernels change physical size over declination — that is, their size in pixels stays fixed as the physical sizes of the pixels themselves change. In principle, then, the wavelets could mix physical scales. In practice, however, the effect of this is limited: the map depth morphology undergoes a smooth transition from small to large scales (again, see figure 5). We should also note that the Fourier space wavelets would encounter regularity issues at the poles; however, this too is not relevant for the ACT footprint, whose largest declination is $-63°$. Fourier space aside, again we can heavily borrow from section 4.3.1 and section 4.3.2 in what follows.

4.4.1 Model estimation

Again, our goal is to construct a sparse representation of the noise covariance matrix. We give an abridged version of the model estimation algorithm from the wavelet model in section 4.3.1, since the steps are highly similar. We take care, however, to explain notable differences in the implementation, especially in relation to the design of the directional wavelets.

1. We apply the same stationarity filter as in section 4.3.1, step 2: flattening the noise power spectra as measured in $\boldsymbol{\mu}_{\mathrm{est}}$. Again, this helps remove the absolute dependence of the noise variance on angular scale, ℓ. We still apply this filter to the maps in harmonic space:

$$\mathbf{C}_{\mathrm{est}}^{-\frac{1}{2}}\mathbf{Y}^{\dagger}\mathbf{W}\boldsymbol{\nu}_i. \tag{4.33}$$

2. In preparation for the wavelet decomposition, we need to transform the maps from harmonic space to 2D Fourier space. To do so, we first return to map space, and then take the DFT to Fourier space:

$$\mathbf{FYC}_{\mathrm{est}}^{-\frac{1}{2}}\mathbf{Y}^{\dagger}\mathbf{W}\boldsymbol{\nu}_i. \tag{4.34}$$

We note that this Fourier space noise vector still contains $6N_{\mathrm{pix}}$ elements, since the DFT, $\mathbf{F}$, is a square matrix.

---

[24]For harmonic space analogs, see e.g., the "curvelets" of [16] and the "steerable wavelets" of [59]. We note that the latter wavelets were used in the scale-discretised, directional wavelet ILC (SILC) analysis of [74]. The total data covariance in SILC is the directional generalization of the NILC covariance, just as our directional wavelet model generalizes our isotropic model.

3. As in section 4.3.1, step 3, our goal is to isolate stationary regions in the transformed noise. Now, we do this by breaking the noise into a set of wavelets defined in Fourier space. We construct the wavelets to be separable radially and azimuthally in the Fourier plane. Modes with comparable radii, $k \equiv |\mathbf{k}|$, map to comparable angular scales, ℓ, as well. For instance, had we constructed the isotropic wavelets from section 4.3 in 2D Fourier space instead of harmonic space, they would have appeared as concentric annuli. Thus, for the radial part of the directional wavelets we reuse the scale-discrete wavelet profiles of [52, 85], only now with a scaling parameter $\lambda = 1.6$ (i.e., we have *increased* the logarithmic wavelet spacing), $J_0 = 5$, and $J = 19$, yielding 16 profiles (see the top panel of figure 12). Compared to the isotropic model, the directional model has fewer and wider radial wavelet kernels. Also note, since we are in Fourier space, we consider the wavelets to be functions of k instead of ℓ. By augmenting the wavelets with an azimuthal part, we create wavelets local not only in their angular scale, k, but in their direction on the sky as well. The azimuthal profile shapes (middle and bottom panels of figure 12) are defined in appendix E; their outer product with the radial profiles yields $M \sim \mathcal{O}(300)$ 2D Fourier space wavelets. The full 2D wavelets are shown in figure 13. As before, these 2D Fourier-space wavelet profiles smoothly overlap to prevent ringing and increase their map-space resolution.

The forward tiling transform associated with the directional wavelets is similar to that of section 4.3. We note that the action of "cropping" out each wavelet from the full 2D Fourier space is akin to the ℓ_{max}-bandlimiting from the isotropic wavelet model, except that we are restricted to a rectangular geometry (as in the tiled model). That is, rather than retaining all harmonic modes up to the wavelet ℓ_{max}, we retain all Fourier modes up the wavelet $|k_{y,\mathrm{max}}|$ and $|k_{x,\mathrm{max}}|$. Thus, each projection matrix $\mathbf{P}_j$ is a $N_{\mathrm{pix,wav}} \times N_{\mathrm{pix}}$ cropping matrix, and $\boldsymbol{\tau}_{j,f}$ is a $N_{\mathrm{pix,wav}} \times N_{\mathrm{pix,wav}}$ is a square diagonal matrix with the jth wavelet kernel values along the diagonal. The forward tiling transform matrix form is again identical to equation (4.5), yielding the following weighted, transformed, tiled noise vectors:

$$\mathcal{T}_f \mathbf{F} \mathbf{Y} \mathbf{C}_{\mathrm{est}}^{-\frac{1}{2}} \mathbf{Y}^{\dagger} \mathbf{W} \boldsymbol{\nu}_i. \tag{4.35}$$

In figures 12 and 13, we clearly observe what makes these wavelets directional: by isolating a compact range of azimuthal angles in Fourier space, their map space kernels are stripy with a well-defined orientation on the sky. Additionally, those figures demonstrate a critical design principle: the "zero-sum" tradeoff in scale, azimuthal (i.e., directional), and spatial locality. In figure 13, for instance, the left panel shows that at large scales (near the "center" of Fourier space), each wavelet exhibits a high degree of scale locality and low azimuthal locality. Correspondingly, their map space kernels (in the top row of the right panel) are spatially extended overall, but are *proportionally* compact along their orientation. At small scales (far from the "center" of Fourier space), each wavelet exhibits successively less scale locality, and successively more azimuthal locality. Correspondingly, their map space kernels (in the bottom row of the right panel) are spatially compact overall, but are proportionally *extended* along their orientation. This behavior is reminiscent of the "parabolic scaling" in [16], wherein the smaller-scale wavelets become progressively more extended. We discuss the engineering of this tradeoff in more detail in appendix E.

4. We again transform from the space in which the noise is assumed to be stationary (Fourier space) into the space where we take the noise to be diagonal (map space), this
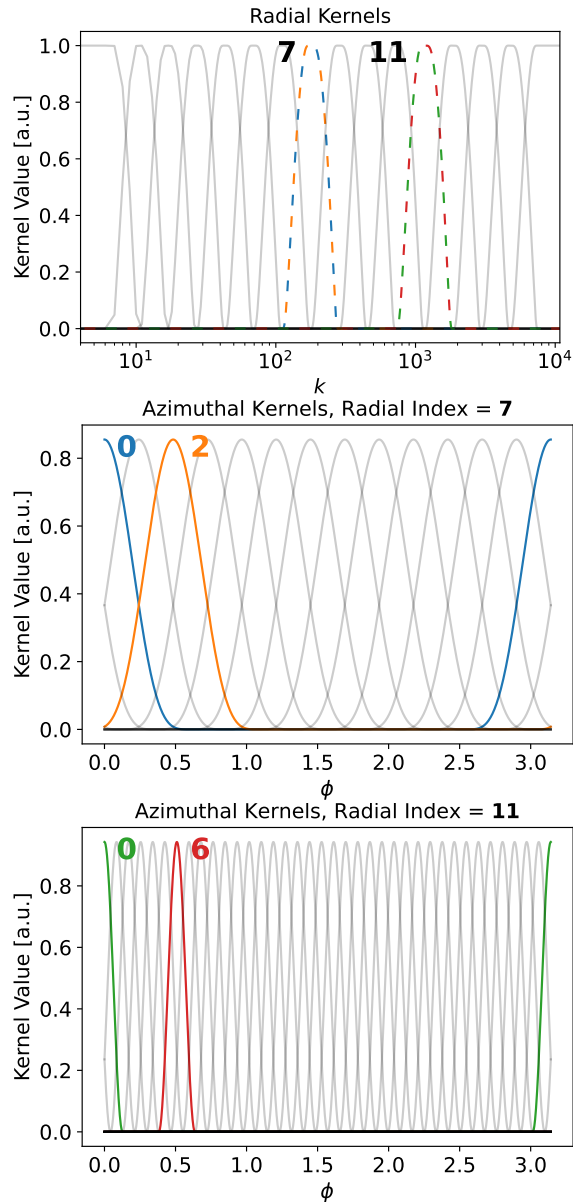
Figure 12. *Top*: radial slices of the Fourier directional wavelet kernels over angular scale. These profiles are again from the scale-discrete wavelet family [52, 85]. The blue and orange dashed wavelet around $k \sim 100$, labeled by a "radial" index of 7, corresponds to the azimuthal slices shown in the middle panel; the green and red dashed wavelet around $\ell \sim 1,000$, labeled by a "radial" index of 11, corresponds to the azimuthal slices shown in the bottom panel. *Middle and Bottom*: we construct a wavelet set with increasing directional locality as a function of k. For example, 13 wavelets tile the azimuthal axis for the lower-k wavelet, while 37 kernels tile the azimuthal axis at the higher-k wavelet. A full 2D wavelet is given by the outer product of the radial and azimuthal slices. Numeric labels are for comparison with figure 13.

time with an inverse DFT. Thus, we assume each pixel in the following vector:

$$\mathbf{u}_i = \mathbf{F}^\dagger \mathcal{T}_f \mathbf{F} \mathbf{Y} \mathbf{C}_{\mathrm{est}}^{-\frac{1}{2}} \mathbf{Y}^\dagger \mathbf{W} \boldsymbol{\nu}_i \tag{4.36}$$

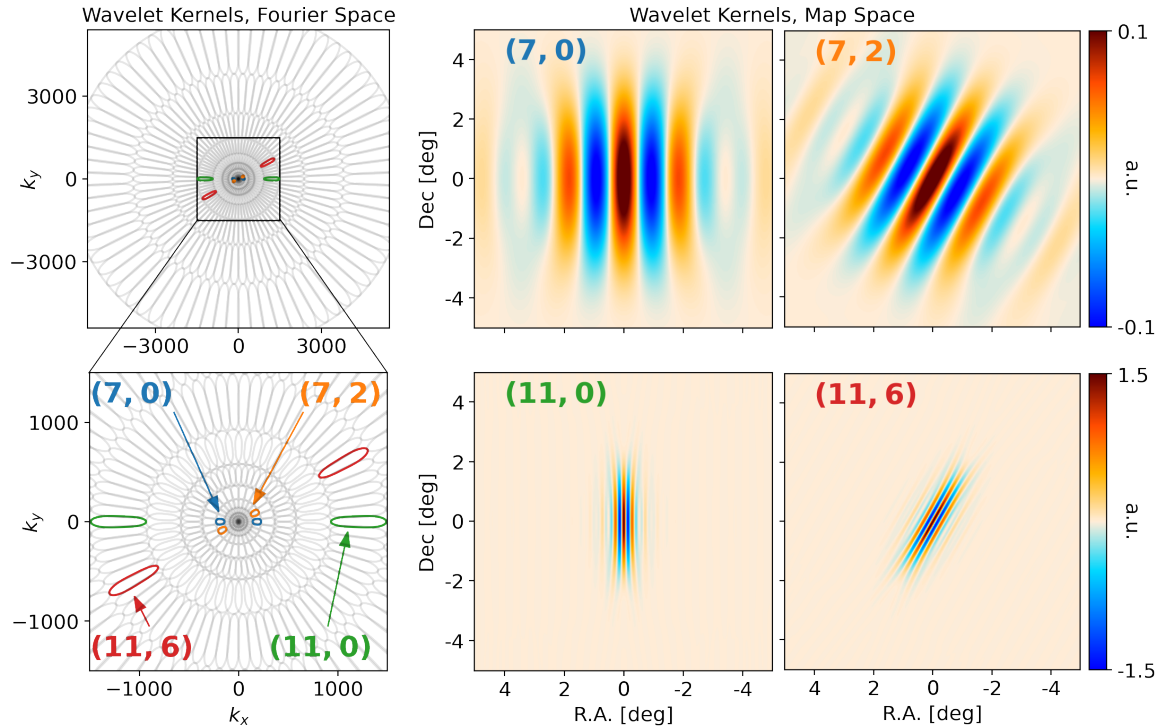is uncorrelated with all other pixels.

Figure 13. *Left*: contour-levels of the full 2D wavelet kernel set. The blue (7, 0), orange (7, 2), green (11, 0), and red (11, 6) 2D wavelets, formed by the product of the 1D profiles in figure 12, are highlighted (numeric labels match those of figure 12). The contours are at the half-maxima of the squared wavelets. We see the wavelets are attuned to "radially oriented" features in 2D Fourier space. *Right*: the map space wavelet kernels for the same blue, orange, green, and red wavelets. Wavelets (7, 0) and (7, 2) pick out low-frequency features by design and are spatially nonlocal. Wavelets (11, 0) and (11, 6) pick out high-frequency features by design and are spatially local overall. The opposite rotation sense in map space versus Fourier space results from the plotting convention (in map space, the R.A. decreases to the right).

5. We evaluate the same pixel-diagonal wavelet power maps, $\mathcal{N}(\mathbf{x})_{i,j}^{ab}$, as in equation (4.25). The only change is due to the rectangular geometry: the normalization factor $f_{j,2}$ is the same form as the tiled model (equation (4.9)), with $N_{\mathrm{pix,wav}}$ replacing $N_{\mathrm{pix,tile}}$. The highly diagonal wavelet covariance matrix, $\mathcal{N}_i$, is likewise the same form as in the isotropic wavelet model.

6. We apply the same smoothing method with the same definition of f as in the isotropic wavelet model (section 4.3.1, step 6) to reduce sample variance.

7. We save to disk the square-root matrices $\mathcal{N}_i^{1/2}$ and $\mathbf{C}_{\mathrm{est}}^{1/2}$.

These steps generate the wavelet model square-root covariance $\mathcal{N}_i^{1/2}$ for the purposes of this paper. We do not show the directional wavelet model maps, but they are qualitatively similar to figure 10. The top panel of figure 11, showing the schematic model estimation steps, is likewise similar for the directional wavelets after substituting spherical harmonic transforms for Fourier transforms.

4.4.2 Simulations and covariance matrix

Apart from the wavelet construction, the process of drawing a noise simulation from the directional model is, again, almost identical to the wavelet model. We give an abridged version of the simulation algorithm here.

1. For each wavelet, draw Gaussian white noise in wavelet map space and take the inner product with the square-root covariance:

$$\mathcal{N}_i^{\frac{1}{2}}\boldsymbol{\eta}_i. \tag{4.37}$$

2. Transform out of wavelet map space and back to wavelet Fourier space with a per-wavelet DFT:

$$\mathbf{F}\mathcal{N}_i^{\frac{1}{2}}\boldsymbol{\eta}_i. \tag{4.38}$$

3. Apply a "backward tiling transform" in Fourier space. That is, again, we need to stitch the wavelets back together to form a single noise Fourier transform. For the same reasons as the isotropic wavelet model, we choose to reuse the forward tiling transform — $\mathcal{T}_b \equiv \mathcal{T}_f$ — noting that we construct our wavelets (appendix E) to satisfy equation (4.16). We therefore have:

$$\mathcal{T}_f^\dagger\mathbf{F}\mathcal{N}_i^{\frac{1}{2}}\boldsymbol{\eta}_i \tag{4.39}$$

where the vector now has the same shape as the Fourier transform of $\boldsymbol{\nu}_i$, and likewise lives in the original, "untiled" Fourier space.

4. We reapply the noise auto- and cross-spectra in harmonic space. To do so, we first transform out of Fourier space and into map space, and then into harmonic space, and then multiply by the square-root noise pseudospectra:

$$\mathbf{C}_{\mathrm{est}}^{\frac{1}{2}}\mathbf{Y}^\dagger\mathbf{W}\mathbf{F}^\dagger\mathcal{T}_f^\dagger\mathbf{F}\mathcal{N}_i^{\frac{1}{2}}\boldsymbol{\eta}_i. \tag{4.40}$$

5. Transforming the entire simulation back into map space completes the noise simulation:

$$\boldsymbol{\xi}_i = \mathbf{Y}\mathbf{C}_{\mathrm{est}}^{\frac{1}{2}}\mathbf{Y}^\dagger\mathbf{W}\mathbf{F}^\dagger\mathcal{T}_f^\dagger\mathbf{F}\mathcal{N}_i^{\frac{1}{2}}\boldsymbol{\eta}_i. \tag{4.41}$$

These steps generate a directional model noise simulation, $\boldsymbol{\xi}_i$, for the purposes of this paper. The schematic depiction in the bottom panel of figure 11 is highly representative of the simulation algorithm. As before, we use equation (4.41) to write the full map-based noise covariance matrix in the directional model:

$$\mathbf{N}_i = \mathbf{Y}\mathbf{C}_{\mathrm{est}}^{\frac{1}{2}}\mathbf{Y}^\dagger\mathbf{W}\mathbf{F}^\dagger\mathcal{T}_f^\dagger\mathbf{F}\mathcal{N}_i\mathbf{F}^\dagger\mathcal{T}_f\mathbf{F}\mathbf{W}\mathbf{Y}\mathbf{C}_{\mathrm{est}}^{\frac{1}{2}}\mathbf{Y}^\dagger \tag{4.42}$$

where we have used that $\mathbf{C}_{\mathrm{est}}^{1/2}$ and $\mathbf{W}$ are Hermitian.

This completes our discussion of the directional wavelet model. By generalizing the wavelets to isolate a particular direction on the sky in addition to a range of angular scales, the directional model naturally captures the spatially-varying noise anisotropy and the scale-dependent map depth simultaneously.
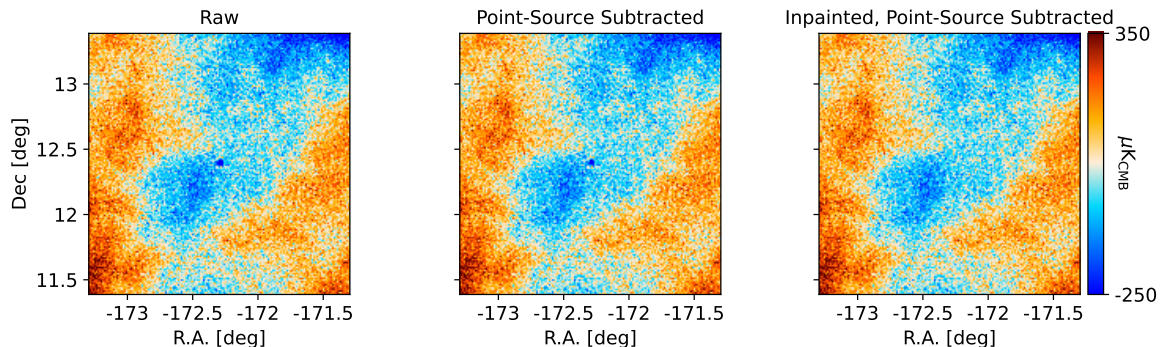
Figure 14. $2° \times 2°$ region centered on M87 in the PA5 f090, first split, T difference map. *Left*: raw difference map (not point-source subtracted). *Center*: using the point-source subtracted maps described in the text partially mitigates the residual point source but does not eliminate it. *Right*: the inpainted difference map.

4.5 Preprocessing and implementation

Now that we have defined the core features and formalism of the models, we complete their discussion with some final implementation details. In sections 4.5.1 and 4.5.2, we describe the data preprocessing we perform on the noise maps, $\boldsymbol{\nu}_i$, prior to using them as inputs to the noise models. This is necessary to mitigate residual map signal that does not cancel in equation (3.3). In section 4.5.3, we define the "pseudospectrum estimate mask", and other masks, used throughout the analysis. Sections 4.5.1, 4.5.2, and 4.5.3 are common to all three models; section 4.5.4 gives further details on the model smoothing step for the wavelet and directional models only.

4.5.1 Source subtraction and inpainting

Each of the preceding models sacrifices some degree of map space locality in order to simultaneously resolve features in harmonic/Fourier space. The raw ACT data, however, contain spatially compact point sources; importantly, the brightest of these sources do not exactly cancel when we construct difference maps in equation (3.3). This is likely due to a variety of underlying causes: many sources are themselves time-variable, or the telescope pointing varies slightly between splits, both of which would leak their signal into the map-based noise as defined in equation (2.1). They may be bright enough that small errors in gain calibration between splits can also induce significant signal-to-noise leakage in equation (3.3). The upshot is that raw difference maps contain spatially compact point-source residuals, which, when left untreated, can bias our noise models.

We mitigate these point-source residuals in two ways. First, we use "point-source subtracted" maps [2, 63]. These are maps in which both time-domain and map-space models of point sources, either from a catalog or matched filter applied directly to the ACT DR6 data, have been fit and subtracted from the raw map products [2, 63]. Use of the point-source subtracted maps as inputs eliminates most of the difference map residuals, but some particularly bright sources remain (see left and center panels of figure 14).

To further increase robustness against point-source residuals, we also inpaint the difference maps in small regions around the sources. For computational simplicity, we do not use a constrained-realization (e.g. [14]) inpainting routine. Rather, we develop a simple, sufficient procedure which leverages the overall smoothness of the noise and compactness of the

point-source residuals. For a given point-source location in the difference map, we inpaint only the pixels within a six arcmin radius — the "inpainted region" — as follows:

1. Set the value of all pixels within the inpainted region equal to the mean value of pixels in the annulus between six and nine arcmin. This removes the point-source residual and replaces it with a reasonable constant value. While this is better than having the large spike from the point-source residual, it will leave a discontinuity in the noise properties at the border of the inpainted region.

2. We would like the large-scale noise features to be continuous across the inpainted region. To achieve this, we first copy the 120 arcmin square region centered on the point source from the *updated* difference map — that is, the map resulting from the first step. We then smooth this copied cutout with a six arcmin FWHM Gaussian kernel. Finally, we adopt the values of this smoothed cutout only in the inpainted region of the difference map. This adds some large-scale modes from outside the inpainted region on top of the constant value from the first step.

3. We would also like the small-scale noise features to be continuous across the inpainted region. For this, we simply add a white-noise realization drawn from the per-pixel difference map variance $(\mathbf{w}_i^{-1})$ to the difference map inpainted region.

These three steps ensure that in a small region around inpainted point-source residuals, the noise features are reasonably continuous. The results of this procedure for a sample point source are shown in the right panel of figure 14. We inpaint around all sources in the compact source catalog described in [71], which contains 1,779 point sources. While only a fraction of this catalog is likely bright enough to cause a noticeable residual, our inpainting procedure is benign and we only inpaint 0.12% (0.28%) of the sky (ACT footprint).

4.5.2   Map downgrading

We typically complete our preprocessing of the difference maps by downgrading their resolution by a factor of two or four, which also reduces their bandlimit proportionately. There is no fundamental reason for this step; rather, we do so only to save disk space and computation time of noise realization ensembles. We resample in Fourier space to perform the downgrading. This has several benefits over pixel-block-averaging. First, the pixel centers in the downgraded pixelization align with pixel centers in the raw pixelization. This is necessary for ensuring SHTs of the downgraded maps are lossless (i.e., the downgraded maps still follow the Clenshaw-Curtis quadrature rule, see [73]). Second, we do not introduce any additional pixel window or aliasing beyond what is already present in the full-resolution maps. Lastly, unlike harmonic space resampling, Fourier space resampling preserves the entire Fourier plane, which is necessary for the directional wavelet decomposition to work. In principle, bandlimiting the difference maps could introduce ringing around bright features, but in practice these features are already suppressed when we take map differences in equation (3.3) and inpaint the difference maps.

4.5.3   Pseudospectrum and boolean masks

Our models also rely on two input masks. The first is an apodized mask that leaves a restricted portion of the sky used to measure the noise pseudospectrum and construct the isotropic filter for the noise models. The second is a boolean mask defining the larger area we model and simulate.
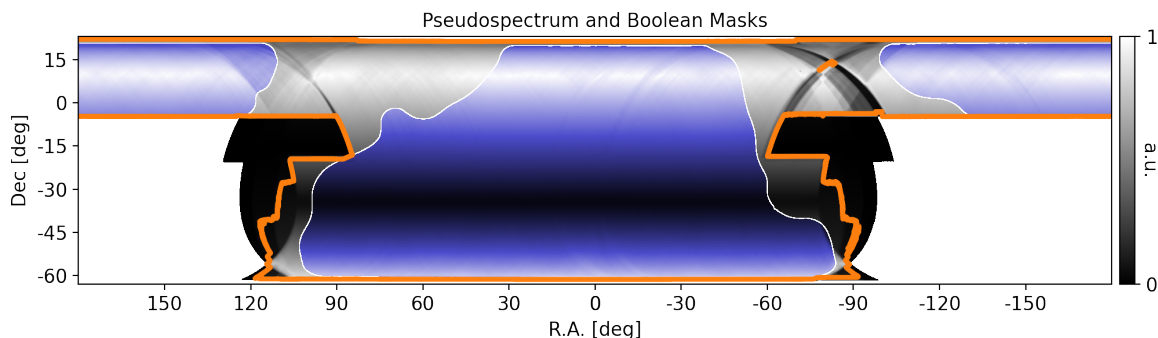
Figure 15. *Grey*: the PA5 f090 cross-linking (square-root of the same data as figure 1. The square-root improves contrast in small values). *Blue*: the region preserved by the apodized pseudospectrum estimate mask, $\boldsymbol{\mu}_{\mathrm{est}}$, overplotted on the cross-linking. *Orange*: the outline of the boolean mask, $\boldsymbol{\mu}_{\mathrm{bool}}$, for PA5.

This first mask is only used to measure the noise pseudospectra which enter the isotropic filter in section 4.2.1 step 2, section 4.3.1 step 2, and section 4.4.1 step 1. The goal of this mask is produce "representative" pseudospectra: downweighting or excluding the noisiest data which would otherwise dominate the spectra, while retaining a reasonable fraction of the ACT footprint. Thus we use a slightly modified version of the nominal ACT DR6 CMB lensing analysis sky mask [71]. This is the intersection of two other masks: a root-mean-square threshold cut on the ACT DR6 f090 and f150 coadd maps of $70 \, \mu$K-arcmin, and a *Planck* 60% Galactic mask [64]. We modify it by then cutting pixels not observed by all splits of all detector arrays. This ensures we can use this one mask version for all of our noise models. Finally, we apodize the mask by three degrees with a cosine profile. The outline of the resulting pseudospectrum mask, $\boldsymbol{\mu}_{\mathrm{est}}$, is shown as the blue region in figure 15, overlayed on top of a cross-linking map. To maximize noise model accuracy one might be tempted to select the same mask here as is used in any downstream analysis, but the ensuing proliferation of noise models could be cumbersome from a data management perspective. Instead, we note that our noise models are reasonably robust to $\mathcal{O}(1)$ variations about the noise spectra as measured in our pseudospectrum mask (see appendix H).[25]

The boolean mask is used to define which data enter our noise models and are simulated. Ideally, we would include all observed pixels in a given difference map, such that we model and simulate all the available data. In practice, however, we found it improved performance to prohibit pixels with the lowest cross-linking from entering the noise models. We detail the motivation and construction of this boolean mask, $\boldsymbol{\mu}_{\mathrm{bool}}$, in appendix F. Once defined, we apply it directly to the difference maps (via $\boldsymbol{\mu}_{\mathrm{bool}} * \boldsymbol{\nu}_i$) before the model estimation algorithms, and directly to the simulations (via $\boldsymbol{\mu}_{\mathrm{bool}} * \boldsymbol{\xi}_i$) after the simulation algorithms. The outline of this boolean mask is shown as the orange curve in figure 15. We see it excludes the regions at the edge of the ACT footprint with practically no cross-linking. Importantly, the boolean mask does not cut any data used in the main DR6 cosmology analyses (the CMB power spectrum and CMB lensing).

---

[25]Furthermore, matching the pseudospectrum estimate mask to some desired analysis region is not guaranteed to achieve optimal agreement between the noise power spectra of simulations and data in that region.

4.5.4 Model smoothing

The preceding noise models are *themselves* noisy approximations of an underlying, true covariance matrix. This statistical "model scatter" is made especially acute by the fact that we build models using a single realization of the input data — a model for each split map. Absent any suppression of the statistical scatter in the estimated covariance matrices, the models will be biased by over-fitting to the data: random fluctuations in the estimated covariance matrix will be recorded as part of the true covariance matrix. Of course, the preceding algorithms do address this issue by smoothing the tiled 2D Fourier power spectra (section 4.2.1, step 6) or wavelet power maps (section 4.3.1, step 6 and section 4.4.1, step 6). These smoothing steps are essentially averages over adjacent modes or pixels in each tiled or wavelet covariance (we discuss whether also averaging noise models over data splits is a viable alternative in section 6). While necessary to avoid over-fitting to the data, however, over-smoothing the noise models may remove structure in the true noise covariance. For example, in the wavelet power maps of figure 10, a given "blob" of high variance might be a statistical fluctuation, such that we *would* want to smooth it away, or might be a real feature in the true noise covariance, such that we *would not* want to smooth it away. Here, we attempt to balance these competing potential biases.

Starting with the wavelet models, we need to specify the function $f(\ell)$ used in section 4.3.1, step 6 and section 4.4.1, step 6. Recall, in those smoothing steps we take each wavelet power map, corresponding to a wavelet with a given $\ell_{\max}$, and smooth it with a Gaussian kernel whose FWHM in map pixels is given by $f(\ell_{\max})$. Thus, we want $f(\ell)$ to lie in the "Goldilocks zone" of not too little, but not too much, smoothing.

Fortunately, we have *a priori* insight into what the noise power maps should resemble: at high-ℓ, the mapmaker inverse-variance maps, and at low-ℓ, the mapmaker cross-linking maps (recall figures 1 and 5).[26] We use this knowledge to estimate the optimal smoothing width, f_{opt}, at a given $\ell_{\max}$, according to the following procedure:

1. Take the raw mapmaker inverse-variance, initially with a Nyquist bandlimit $\ell_{\max} = 21,600$, to be the "true" noise covariance. Draw a white-noise realization from this covariance to represent a single realization of the map-based noise in the data.

2. With varying amounts of smoothing, construct white-noise models using this realization of the map-based noise as the input. This is done by simply squaring the observed noise in each pixel, and smoothing with a Gaussian kernel whose FWHM is f pixels.

3. Compare the agreement between the "estimated" white-noise covariance in the models with the original, "true" noise covariance. The optimal smoothing, f_{opt}, corresponds to the model that maximizes this agreement.

4. Repeat these steps by varying what is taken to be the true noise covariance, by downgrading the inverse-variance map to a lower $\ell_{\max}$ and/or including the cross-linking map. Find f_{opt} in each case.

The fourth step above is important to ensure the fiducial input covariance is as realistic as possible. First, the wavelet power maps probe noise structure at different resolutions, as set by their $\ell_{\max}$. Consider the limiting case of a white-noise covariance defined by the

---

[26]To be precise, these two mapmaker products trace the spatial variation of the map depth. It is their inverse that traces the map noise power.
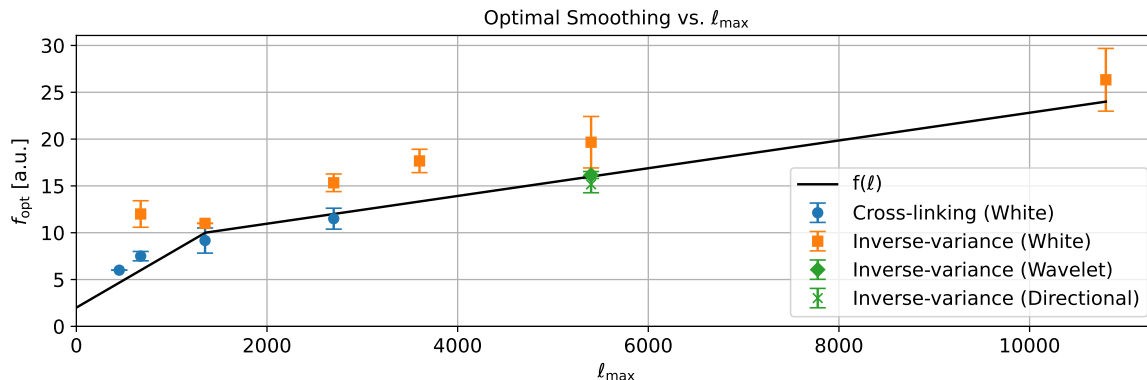
Figure 16. Optimal smoothing factors, f_{opt}, as a function of bandlimit, using the procedure in the text (also see appendix G). The blue circles are obtained by comparing a simple white-noise model to a cross-linking plus inverse-variance-like noise model. The orange circles correspond to a white-noise model following the inverse-variance only. The green diamond (cross) are obtained by comparing the wavelet (directional) noise model to an inverse-variance-like noise covariance. Each data point shows the mean and 1σ range over the six ACT DR6 arrays/frequencies (first split of each). The solid black line shows our chosen form of $f(\ell)$ for use in the actual model estimation of section 4.3.1 and section 4.4.1.

mapmaker inverse-variance map. In this case, even though there is a "true" noise covariance at full resolution, such a covariance has different spatial structure when downgraded to lower resolutions. Thus, the optimal smoothing level at those downgraded resolutions may not be the same as at full resolution. Second, we observed in figure 5 that the spatial noise structure resembles the mapmaker cross-linking at large scales. Thus, by including the (downgraded) cross-linking in our input covariance at lower values of ℓ_{max}, the resulting optimal smoothing level is appropriate for models of the actual data.

This procedure balances the tension between preserving the true covariance structure and suppressing random fluctuations in realistic wavelet power maps. We show the results in figure 16.[27] In general, the optimal smoothing factor, f_{opt}, increases with ℓ_{max}. At large scales — that is, scales dominated by correlated noise — the cross-linking governs the spatial noise structure, and $f_{\text{opt}} \sim 5-10$. At small scales, where the inverse-variance dominates the spatial noise structure, f_{opt} increases toward $\sim 20-25$. The open green diamond and cross in figure 16 correct for the fact that our wavelet noise models are not simple white-noise models (also explained further in appendix G). They indicate the optimal smoothing for the wavelet noise models is slightly less than implied by our simple white-noise procedure.

In light of these results, we opt for a simple ad-hoc form of $f(\ell_{\text{max}})$ in all arrays and frequencies, using the model shown in figure 16.[28] This respects all arrays' preference for low smoothing factors at low ℓ, with a shallow increase toward larger smoothing factors at large ℓ. This curve also remains $\sim 10-30\%$ below the white-noise results for compatibility with our wavelet models. We also see no significant effect in the results of section 5.1 when varying $f(\ell_{\text{max}})$ by ~ 5 at all but the largest scales. Therefore, in setting the form of $f(\ell_{\text{max}})$ we are primarily concerned with capturing the overall trend than specific, minor variations between arrays. We plan to improve our smoothing procedure in future work.

---

[27]We give the detailed version of this approach and define a quantitative measure of "model-to-truth" agreement in appendix G.

[28]The lines from $(0, 2)$ to $(1,350, 10)$, and then through $(5,400, 16)$ in $(\ell, f(\ell))$ space.

For the tiled model, there is no mapmaker product that gives an *a priori* understanding of what the 2D Fourier power spectra should look like. Thus, as we discussed in section 4.2.1, step 6, we simply adopt the same smoothing scale as was used in [57] and [63].

5 Model evaluation and results

Having constructed the tiled, isotropic wavelet, and directional wavelet noise models, we examine their ability to reproduce ACT map-based noise statistics. As discussed in the previous section, we do so by averaging over large ensembles of noise simulations. A data difference map and example noise simulations for each model are shown in appendix A. For ease of computation, we have chosen to show results from models bandlimited to $\ell_{\max} = 5,400$, and averages of 30 simulations.

Because of the density of the data and model noise covariance in map space, a direct comparison of those matrices is not possible. As an informative substitute, we instead examine the noise properties of section 3.2 as captured in simulations. Thus, in section 5.1, we compare simulations to the data by: (a) isotropic noise auto- and cross-spectra (large-scale noise correlations), (b) 2D Fourier power spectra (spatially-varying noise anisotropy), and (c) effective "inverse-variance" at different angular scales (scale-dependent map depth). Each comparison is analogous to viewing the full map-based noise covariance matrix along a different "slice", or through a different "lens": it visualizes a particular salient aspect of the matrix structure, since doing so for the entire matrix at once is difficult. In section 5.2, we apply the noise simulations to a realistic analysis pipeline. In particular, we examine the covariance matrix of the ACT DR6 power spectrum bandpowers — that is, the determination of the errors on the power spectrum. We compare a common analytic prescription for the covariance matrix and a Monte Carlo estimate from simulations alone.

5.1 Noise covariance model-data comparisons

We examine the noise properties from section 3.2 as captured by our simulations and compare to the data. Thus, we directly probe the extent to which each model achieves its design goals. As discussed, each property visualizes a slice of the full covariance matrix, either constructed from the *single* available realization of the data, or averaged over an *ensemble* of simulations. If we have built a good noise model, the covariance slice measured from the data should be consistent with a single noisy realization of the ensemble average covariance slice. Said another way, the ensemble average slice is each model's postulated long-run average of the data slice, if it were possible to rerun the experiment many times. Thus, for a successful model, we expect the data slice to "scatter" around the smoother long-run average provided by the simulations, as is the case in a typical goodness-of-fit analysis. Indeed, in future work, we aim to add a full χ^2 consistency test of data and simulations, which would require application of the full inverse noise covariance matrix, $\mathbf{N}_i^{-1}$, to map vectors (see section 6 for a discussion of the inverse covariance matrix). At any rate, we should keep this signature of good model performance in mind in what follows.

5.1.1 Large-scale noise correlations and correlated frequencies

Extended atmospheric signals produce most of the non-trivial map-based noise properties, including pixel-pixel correlations, and correlations between frequencies on the same detector array. Figure 17 demonstrates the noise models' performance at reconstructing the noise auto-spectra and cross-spectra between frequencies. Generally, all models perform similarly
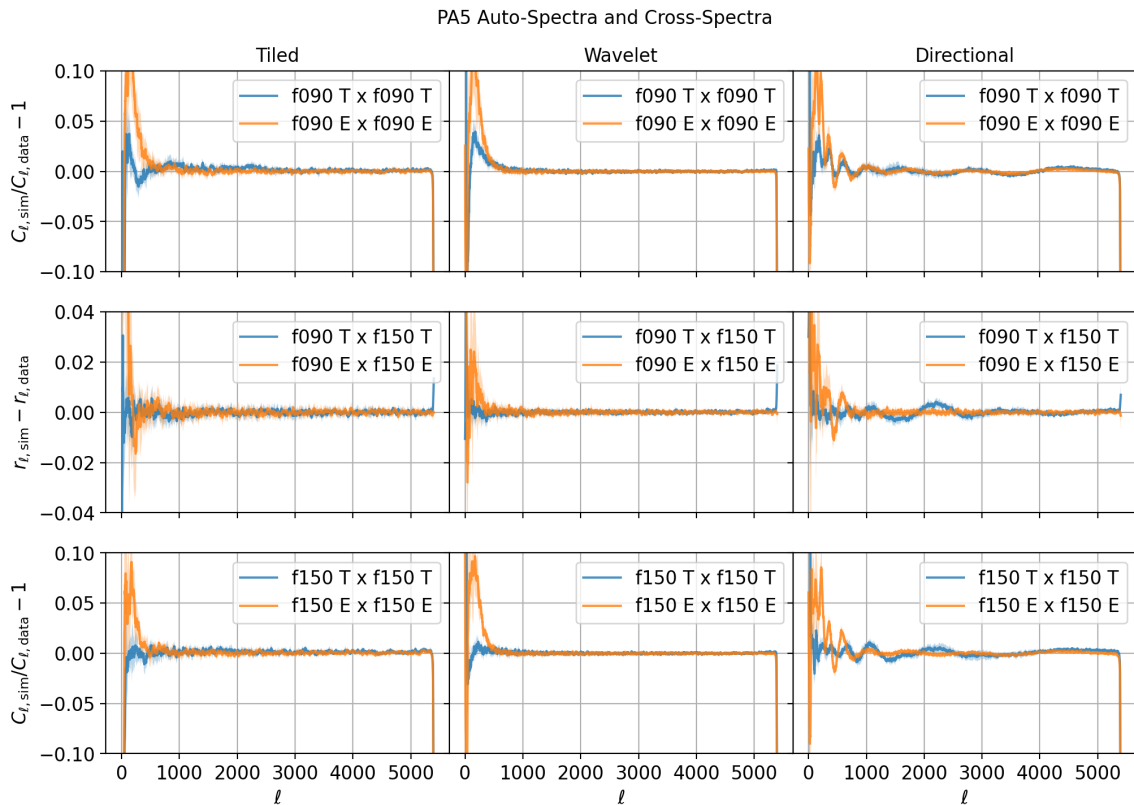
Figure 17. Comparisons of PA5 noise auto- and cross-spectra between the simulations and the data. The solid lines (shaded bands) are the means (1σ range) over the eight splits. The top row shows f090 power spectrum ratios. The bottom row shows f150 power spectrum ratios. The middle row shows f090 $\times$ f150 correlation spectra differences, calculated according to equation (3.9). For all three rows, the nominal result is a value of 0. All spectra are measured in the same apodized pseudospectrum mask from section 4.5.3 and are smoothed with a tophat kernel of size $\Delta\ell = 25$. Recall the models are bandlimited to $\ell_{\max} = 5,400$. Temperature spectra perform well at all scales, exhibiting no more than few-percent deviations at low-ℓ. Polarization spectra appear robust for $\ell \gtrsim 300$, but exhibit $\sim 10\%$ deviations for $\ell \lesssim 300$.

well in reconstructing the noise TT, EE auto- and TE cross-spectra. Their mean values are within 1% of nominal (zero for the correlation spectra differences and the auto-spectra ratios, since in figure 17 we have subtracted 1 from the ratios). The directional model exhibits oscillations, likely due to its Fourier-space wavelets, but they are at or below the percent-level. As discussed in section 4.4, these wavelets change physical size over declination, leading to scale-mixing in harmonic space. We note that the oscillations roughly align with the wavelet kernel radial profiles in the top panel of figure 12, hinting at the wavelets as the underlying cause.

For all three models, large departures from nominal are confined to large scales ($\ell \lesssim 300$), especially in polarization. One possible explanation for this behavior is the pseudospectrum filter, which is applied to all of our models. This filter places a lower bound on the map space locality of features in the simulated noise. As we show in section 6, this lower bound happens to be about three times higher in polarization than temperature. In other words, it is more difficult to contain locally noisier regions of the sky in polarization, and we would interpret
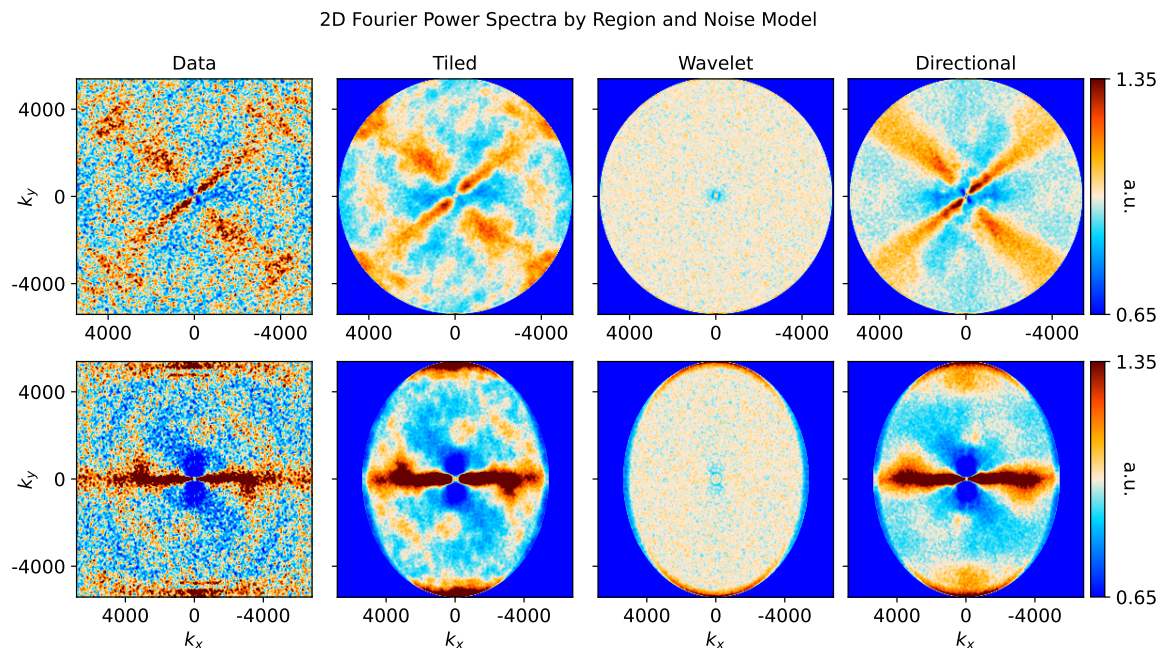
Figure 18. The same 2D Fourier power spectra treatment as figure 4, for both data and noise simulations. The top (bottom) row corresponds to the same regions as the left (right) panel of figure 4. The first/second/third/fourth columns correspond to the data/tiled/wavelet/directional models. The model bandlimit ($\ell_{\max} = 5,400$) causes the oval power loss in the simulations. The bandlimit is nearly circular in 2D Fourier space in the top row, since this is from a region near the equator. The bandlimit becomes "squished" in the x-direction in the bottom row, since this region is at a higher declination. That is, the pixels in the x-direction are physically closer together at these latitudes, and so the 2D Fourier space samples higher frequencies. Note we account for this in our definition of the 2D Fourier coordinates. The tiled and directional models capture the main anisotropies observed in the data, with the tiled model resolving more fine features. The wavelet model has no ability to measure anisotropies by construction.

figure 17 as showing some noisy, large-scale polarized regions bleeding into the analysis mask. We note, however, that ACT DR4 power spectra were limited to $\ell_{\min} = 350$ in polarization and $\ell_{\min} = 600$ in temperature [19]. Overall, we take the results of figure 17 as a marker of good noise model performance across all three model types.

5.1.2 Spatially-varying noise anisotropy

While the tiled and directional models are designed to capture the location-dependent stripiness in the ACT noise, the wavelet model is not. To evaluate their performance in this regard, we simply repeat the analysis in figure 4, substituting an average over simulations for the data. The success, or lack thereof, in each model is visually apparent in figure 18. Both the tiled and directional models succeed at reconstructing the primary noise stripiness in each region: the "x-like" pattern near the equator, and the horizontal "bar" in the poorly cross-linked region. The tiled model appears to capture some secondary features in the data noise that are not oriented radially; recall the directional kernels (figure 13) are primarily designed around radial features. The wavelet model stands out: by construction, it has no ability to measure local noise anisotropy. We also observe that the high frequency "covariance scatter" in the *data* 2D power spectrum is substantially suppressed in the *simulation* spectra. This is

expected: as discussed, in the simulation spectra we visualize the true, underlying covariance structure postulated by each model, of which the data spectrum is a single realization. This also inspires confidence in the model smoothing from section 4.5.4: the models are not obviously over-fit to the random fluctuations in the data covariance. In sum, the tiled model and directional models succeed at capturing the primary features of the spatially-varying noise anisotropy in the data, while the wavelet model captures no such features.

5.1.3   Scale-dependent map depth

In addition to spatial correlations and stripiness, we showed how the atmosphere and scan strategy together induce a scale dependence to the map depth in section 3.2. In particular, we showed that the small-scale noise power appears to follow the morphology of the mapmaker inverse-variance maps, while the large-scale noise power is sensitive to the scan cross-linking. We compare the ability of the noise models to reproduce these effects in figures 19 and 20. When filtered for small scales (figure 19), all three models succeed at reproducing the mapmaker inverse-variance-like morphology in the data. When filtered for large scales (figure 20), the two wavelet-based models succeed at reconstructing the mapmaker cross-linking-like morphology, for which they were nominally designed.

The tiled model, while not explicitly designed for this purpose, also performs moderately well. This is because a scale-dependent map depth is equivalent to a spatially-dependent noise power spectrum. The tiles do model a spatially dependent (2D Fourier) power spectrum, but limited to a resolution of the tile size ($\sim 4°$). As a result, some features in the tiled large-scale spatial noise power are not resolved — for instance, the fine structures near a declination of $-20°$. Other structures are "squared-off", such as the transitions between high- and low-cross-linking regions.

For the large-scale maps, we also note that all three models show slightly less inverse-variance (more noise power) than the data. This is a restatement of the result in figure 17 that the polarized noise power spectra ratios are elevated for $\ell \lesssim 300$. Finally, as for the 2D Fourier spectra of figure 18, we observe the same suppression of high-frequency scatter in the simulated averages as compared to the data. This is again expected, and further supports the model smoothing of section 4.5.4. In sum, the two wavelet models perform somewhat better than the tiled model at capturing the scale-dependent map depth in the data.

In short, the models behave as designed. All three models reconstruct the overall noise auto- and cross-spectra. The tiled model is best at capturing spatially-varying noise stripiness; it performs less well at capturing scale-dependent map depth. The wavelet model cannot capture stripiness; however, it is best at capturing scale-dependent map depth. Finally, the directional model performs well in both analyses.

5.2   ACT DR6 power spectrum covariance

In this section, we examine the implications of our noise simulations on the estimation of the covariance matrix for the CMB angular power spectrum. Specifically, we examine the matrix, Ξ, given by:

$$\Xi_{b,b'}^{WXYZ} \equiv \langle \Delta C_b^{WX} \Delta C_{b'}^{YZ} \rangle \tag{5.1}$$

where ΔC_b^{WX} is the *residual* of the recovered power spectrum in ℓ-bin b, between fields W and X (one of T, E, or B), with respect to a fiducial model. An accurate covariance matrix is a required deliverable of any CMB experiment: inaccuracies would directly propagate into cosmological parameter uncertainties and could also skew power spectrum pipeline null tests.
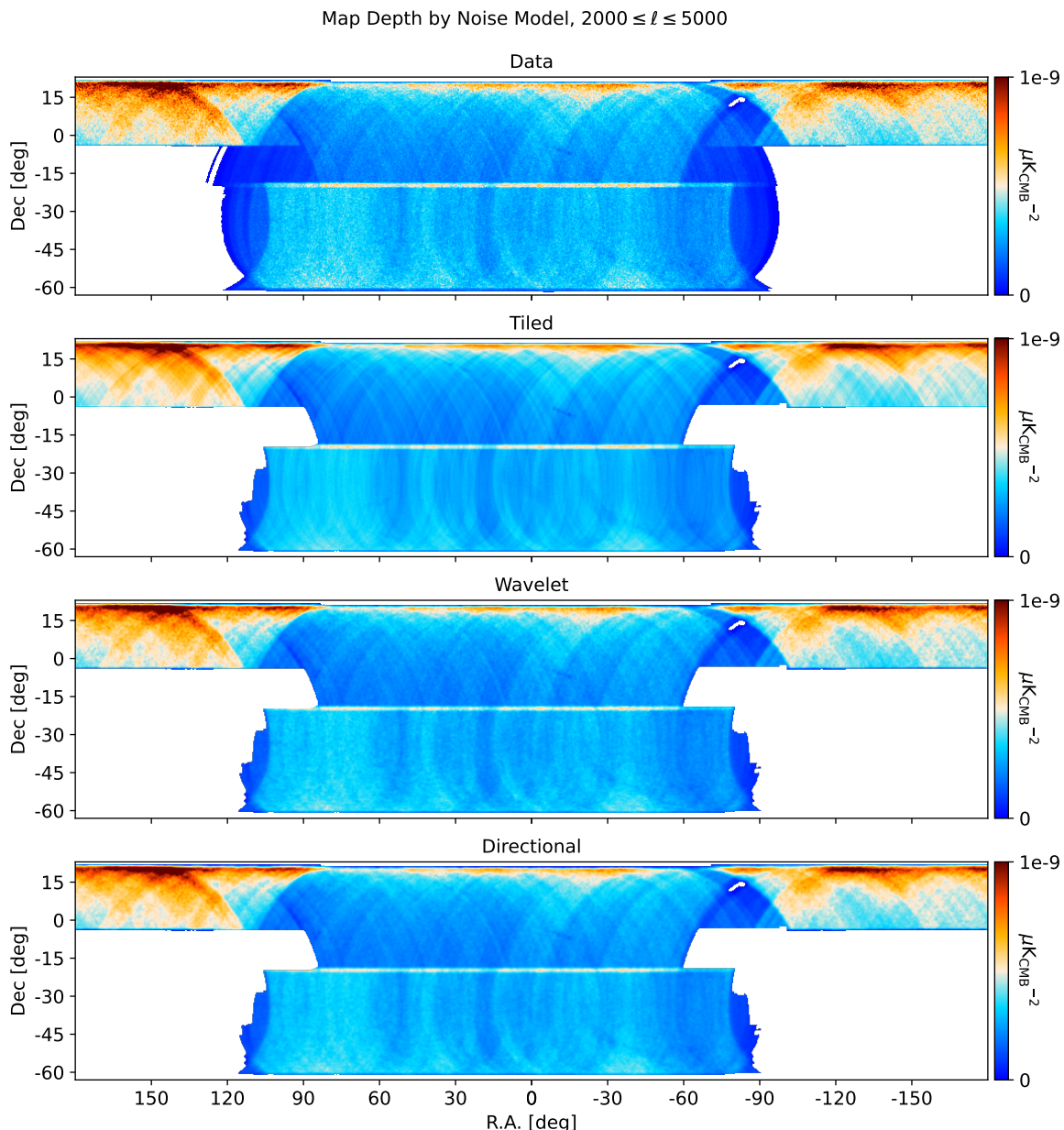
Figure 19. The same small-scale "inverse-variance map" as the top panel of figure 5, for both data and noise simulations. The first/second/third/fourth rows depict the data/tiled/wavelet/directional models. As is discussed in section 4.5.3, we see that simulations are masked with the boolean mask that removes sky regions with the lowest cross-linking. All models do a good job reproducing the high-ℓ map depth in the data.

While the exact calculation of equation (5.1) is difficult (see e.g. [68]), several approximations exist. These incorporate knowledge of a fiducial sky model, instrumental noise, and incomplete sky coverage into computationally tractable, *analytic* expressions of the covariance matrix (e.g. [13, 15, 22, 31, 38, 54]). These analytic expressions may also be augmented with simulations to correct for some of their assumptions, for example, by rescaling the covariance matrix at low-ℓ to account for masking point sources with small holes [53, 66, 68].
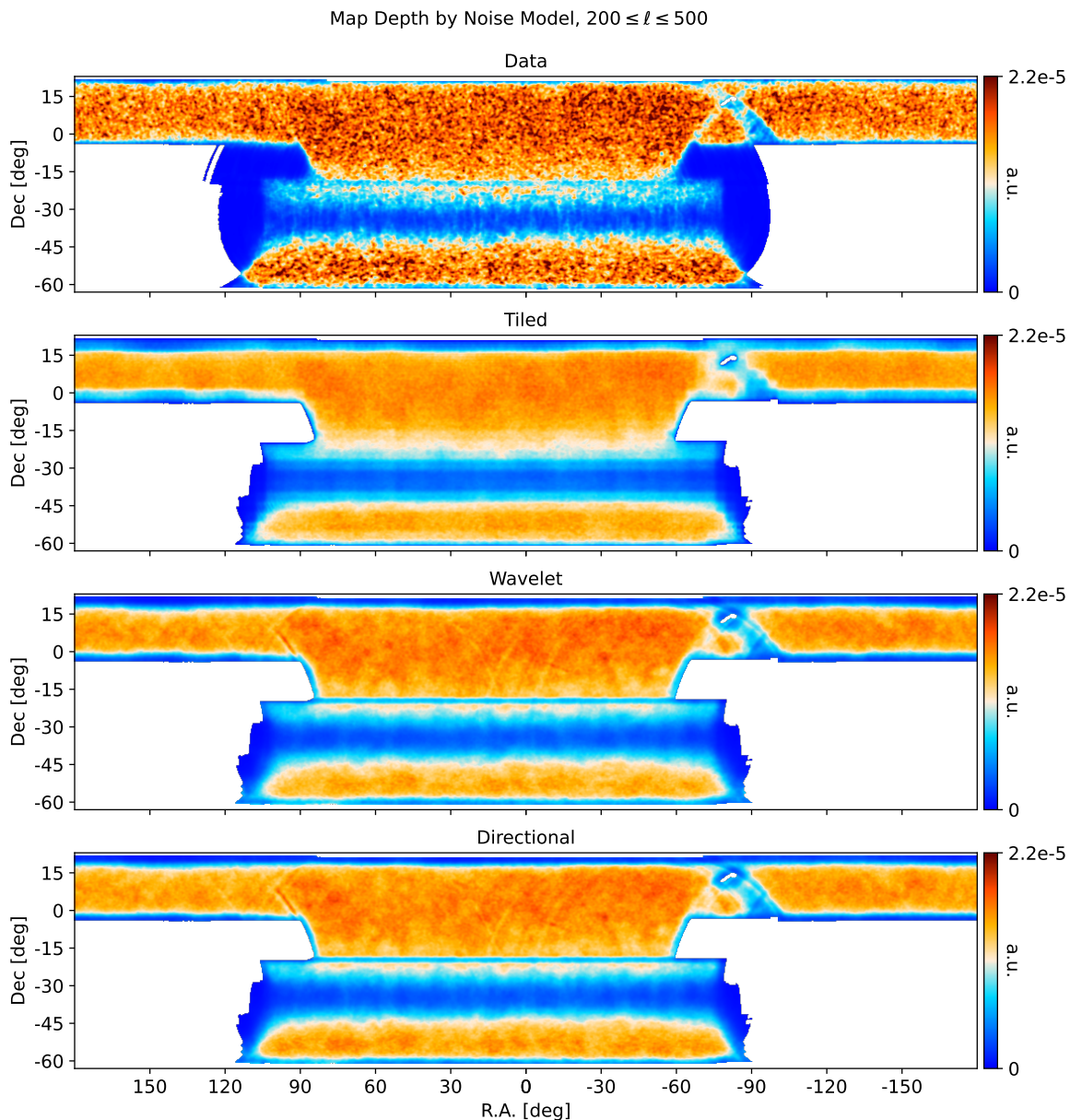
Figure 20. The same large-scale "inverse-variance map" as the bottom panel of figure 5, for both data and noise simulations. The first/second/third/fourth rows depict the data/tiled/wavelet/directional models. All three models exhibit slightly less inverse-variance (more noise power) than the data, as is expected from the elevated polarized power spectra ratios in figure 17 for $\ell \lesssim 300$. The wavelet and directional models capture fine structures in the data; the tiled model resolution at large scales is limited to the tile size — map features are "squared-off" or unresolved.

As far as their treatment of instrumental noise, these semi-analytic methods adopt one of two approaches. The first approach assumes the noise is diagonal in the harmonic domain — that is, completely described by the noise power spectrum. This is the default approach in the `NaMaster`[29] [38] and `pspy`[30] [54] codes, which together form the SO power spectrum

---

[29]https://github.com/LSSTDESC/NaMaster.

[30]https://github.com/simonsobs/pspy.

pipeline, `PSpipe`.[31] The second approach generalizes the first by combining knowledge of the noise power spectrum and the inhomogeneous white-noise depth (e.g. [32, 68]).

In the following subsections we compare equation (5.1) estimated from an analytic expression following the former approach (assuming diagonal noise in the harmonic domain), to a Monte Carlo estimate from simulations alone. We will examine analytic expressions following the latter approach in future work. We focus on the PA6 f150 band, whose results are representative of the full DR6 data.

5.2.1 Analytic and Monte Carlo covariances

We construct an analytic covariance matrix assuming the signal and noise are completely described by their respective power spectra — that is, assuming uniform white-noise depth. In particular, we use the methods of `pspy` available in `PSpipe`. Our implementation adopts the best-fit ΛCDM cosmology from [67] as the primary CMB power spectrum signal, and adds a foreground model power spectrum consistent with [29], derived from maps with a 15 mJy source flux cut. We incorporate the noise pseudospectra estimated directly from the DR6 data (i.e., the unnormalized version of the spectra in figure 3), following equation 8 of [54] (see also equation 2.55 of [38] and equation 10 of [53]). To keep our analysis straightforward and avoid issues with point source holes, we reuse the pseudospectrum estimate mask, $\boldsymbol{\mu}_{\mathrm{est}}$, from the previous sections (shown in figure 15).

Next, we prescribe the simulations from which we build a Monte Carlo covariance matrix. We first draw signal map realizations from the same CMB-plus-foreground signal model that entered the analytic matrix, and convolve them with the PA6 f150 beam. We then add noise realizations in map space using the methods of section 4. We evaluate the power spectra of 300 such simulations, for each of the eight PA6 f150 splits, and directly compute the Monte Carlo covariance matrix (equation (5.1)) by averaging the residuals of the estimated power spectra compared to the input signal power spectrum. As a sanity check we also generate a set of "Gaussian" Monte Carlo simulations whose noise realizations are simply drawn from the noise power spectra, and estimate their covariance matrix following the same method. Both the analytic and Monte Carlo matrices account for the mode-coupling induced by the analysis mask,[32] and are binned in ℓ to improve their conditioning.

5.2.2 Comparison

We compare the diagonals of the Monte Carlo and analytic matrices in figure 21. In terms of equation (5.1), that is, we only consider the elements of Ξ with $b = b'$ and $W = X = Y = Z$. As expected, the "Gaussian" simulations produce a Monte Carlo matrix whose diagonal agrees with the analytic expression to a few percent, demonstrating that our pipeline is not significantly biased. We then note that in spite of their different methods, our three noise models deliver remarkably similar results. In particular, the TT (EE) block of the analytic covariance appears to underestimate its diagonal by up to $\sim 15\%$ ($\sim 20\%$) for $\ell \gtrsim 2,000$ $(1,000)$. Such discrepancies exceed what has been found in comparable analyses of *Planck* data. For instance, [22] compared the diagonals of a Monte Carlo covariance estimate and an analytic expression assuming uniform white-noise depth and found errors of at most a few percent. Meanwhile, [53] performed a similar comparison and suggest errors of order

---

[31]https://github.com/simonsobs/PSpipe.

[32]The noise power spectra entering the analytic matrix were measured after applying the same Fourier space ground pickup filter as in ACT DR4 [19] and figure 2. Likewise, we applied this filter to the simulations. The final matrices then deconvolve the transfer function induced by this pickup filter.
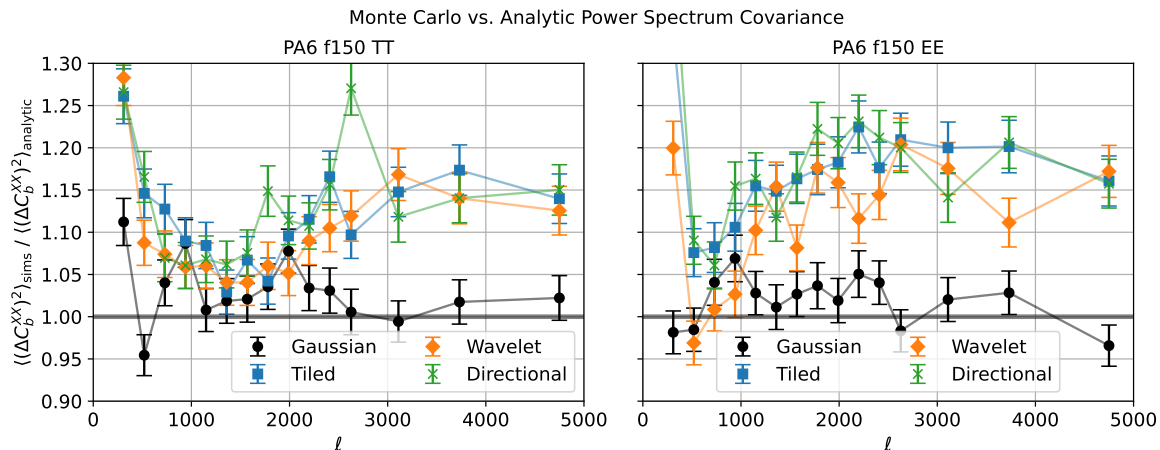
Figure 21. Ratios of the power spectrum covariance matrix diagonal elements between a Monte Carlo construction using 300 noise simulations, and an analytic approach common in the literature. The nominal result is a value of 1. TT (EE) ratios exhibit an excess of up to $\sim 15\%$ ($\sim 20\%$). The "Gaussian" simulations are a sanity check: they are drawn directly from the measured noise power spectra and should match the analytic covariance. Significant deviations of the "Gaussian" simulations from nominal (i.e., in TT for $\ell \lesssim 300$) suggest discounting this comparison at those scales. Data points are plotted at the center of the multipole bins, and show the per-bin mean and 1σ error in the mean coming from the 300 simulations.

$\sim 10\%$. At a minimum, our noise simulations suggest increasing the ACT DR6 power spectra covariance by up to $\sim 20\%$ in polarization.

The agreement of our three noise models in figure 21 hints at a common underlying mechanism. As discussed, one ACT noise property not accounted for in the analytic covariance of this section is the spatially-varying map depth (regardless of whether the map depth is simultaneously scale-dependent). Thus, the natural next step would be to evaluate an analytic expression that includes the inhomogeneous ACT white-noise depth [32], as was used for *Planck* [66, 68], assessing whether it could sufficiently capture the additional variance observed in our simulations. These results will inform the approach to constructing a power spectrum covariance matrix for ACT DR6, and will be explored in future work.

6 Discussion and conclusions

For ACT DR6, the advent of three separate noise models allows downstream analysis pipelines to evaluate results against a range of map-based noise properties. For instance, we showed how propagation of noise simulations through the ACT DR6 power spectrum pipeline can help measure the extent to which analytic covariance matrix expressions accurately capture the nontrivial ACT noise properties (e.g. [32, 53]). Likewise, the ACT DR6 lensing analysis [71] utilizes these models to demonstrate that the lensing inference is robust with respect to the complicated ACT noise. While checking pipelines for robustness against different noise models can be useful, it is too early to say whether one model is the most performant for particular applications. Finally, consumption of large noise simulation ensembles by ACT and community users requires efficient code (see appendix I) and a simple interface to those products. This code — mnms — is made public with this paper, and is out-of-the-box applicable to future ground-based CMB experiments such as SO and CMB-S4. Nevertheless, it is worthwhile to evaluate key model assumptions and suggest directions for future development.

Measuring noise from split differences. Throughout this paper, we have assumed $\boldsymbol{\nu}_i$ to be an accurate noise estimate. In section 3.1 we demonstrated this is a reasonable assumption when averaged over a large sky fraction. However, in regions of especially compact or bright foregrounds, we encounter residual signal in our difference maps. As discussed in section 4.5.1, we mitigate point-source residuals with inpainting. However, we do not handle bright, extended source residuals, which can be prominent in the Galactic plane (see figure 23). The underlying cause of these residuals is likely similar as for point sources: a combination of pointing variation in the telescope and small relative calibration errors between splits. Our noise models will therefore be biased in these regions. Future work could include more sophisticated inpainting than our implementation in section 4.5.1 to handle bright, extended regions.

Non-identically distributed splits. As noted, unlike ACT DR4, we do not assume that the noise in separate splits is identically distributed. Therefore, we do not average noise models over splits in their sparse basis. Whether to do so is tightly coupled to the model smoothing: averaging over the eight splits of ACT DR6 would reduce the amount of required model smoothing. However, if splits contain different noise structure, averaging over them would bias every model. In fact, we know *a priori* that the noise is not identically distributed over splits because their inverse-variance maps vary considerably. We indeed observe percent-level improvement in low-ℓ power spectra ratios when treating splits separately versus when averaging over splits. In principle, including inverse-variance maps as a *mode-decoupling* filter in the wavelet models (as we discussed, and forewent, in section 4.3.1 and section 4.4.1) would eliminate the barrier they present to averaging over splits.

Noise pseudospectrum filtering. Each of our noise models includes the same noise pseudospectrum filter. The benefits of this filter were discussed in section 4 and are explored further in appendix H. However, this filter has a downside: it limits our ability to model high-resolution noise features in map space, especially in polarization. To see how, consider that the noise pseudospectra in the top panel of figure 3 downweight small scales relative to large scales. Thus, the noise pseudospectrum filter has a similar effect in the noise models as the instrumental beam has on a given sky signal: by suppressing small scales it "blurs" compact features. We can see this by examining the "beam-like" real-space profiles corresponding to the transform of noise pseudospectra. Results are shown in figure 22 for the PA5 f090 band. The FWHM of the EE and BB real-space profiles are both $\sim 5°$, whereas the TT profile is more spatially compact, with a FWHM of $\sim 1°$. In other words, given a point-like noise feature in both temperature and polarization, we would expect our simulations to resolve it to $\sim 1°$ in temperature and $\sim 5°$ in polarization. This may explain, for instance, the greater deviations observed in the low-ℓ power spectra ratios (figure 17) for EE than TT: noisy regions from farther outside the analysis mask in polarization than temperature are smoothed into the mask, skewing the ratios. The different real-space widths in figure 22 are driven mainly by the low-ℓ shape of the noise pseudospectra in figure 3: in some bands, the TT noise power at the largest scales ($\ell \lesssim 100$) level off and even decrease. This does not occur in polarization. We will seek to better understand the shape of the low-ℓ noise power spectra in future work.

Powers of noise covariance matrix. In this paper, we only interact with the full noise covariance matrix by drawing samples from the Gaussian distribution it defines. Downstream analyses may wish to raise the matrix to an arbitrary power and then apply it to a map vector.
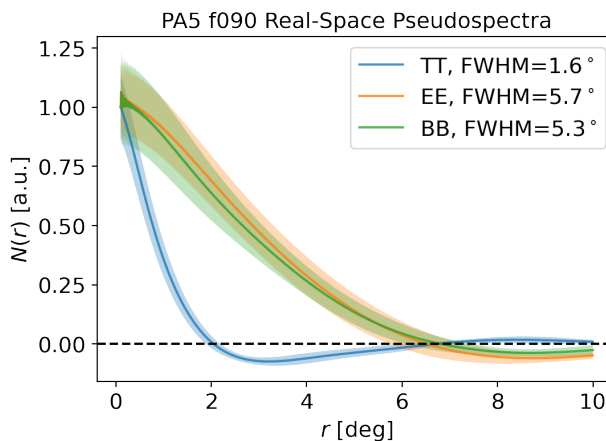
Figure 22. "Beam-like" real-space profiles corresponding to the transform of the noise autospectra from the top panel of figure 3 for PA5 f090. The profiles have been peak-normalized. Solid lines (shaded bands) denote the mean (1σ range) over splits.

In particular, Monte Carlo methods, inverse-covariance filtering, and Wiener filtering require the full inverse noise covariance matrix. Such an implementation is beyond the scope of this work; nevertheless, we can comment on its feasibility. Unfortunately, due to the overlap of tiles in map space (likewise, of wavelets in harmonic/Fourier space) the class of noise covariance matrices prescribed here are not analytically invertible. An exception would be the case of non-overlapping (i.e., tophat) tiles or wavelets, but with a detrimental impact to their performance as noise models (as noted in the text) from their lack of smoothness. In principle, the exact application of the inverse noise covariance matrix to a map vector is still possible with a conjugate gradient method, but may be slow. We will explore this and other alternatives in future work.

Prospective. These results offer concrete insights for the design of next-generation, ground-based CMB experiments. Foremost among them is the central role of the map cross-linking in modulating large-scale noise power. Combined with the tight relationship between cross-linking and the noise anisotropy pattern, spatial uniformity of cross-linking is as, if not more, important to realizing a successful map-based noise model than uniformity of detector hits. We acknowledge that achieving uniform cross-linking may encounter more observational constraints than uniform detector hits (e.g., the dynamics of the telescope and receiver). In a similar vein, projects seeking to use these noise modeling methods should design scan strategies that facilitate identically distributed map splits. While formally difficult, maximizing the uniformity of heuristics across splits (again, e.g., cross-linking, inverse-variance) would go a long way toward allowing noise models to be averaged over splits. Projects should also seek to avoid unnecessary "hard edges" in their scan strategies. This includes not only the observation footprint boundary, but also adjacent regions of disparate cross-linking or inverse-variance levels. Such boundaries within the ACT DR6 footprint proved especially challenging to handle from a noise modeling perspective.

Finally, we emphasize the utility of a fast and flexible noise model implementation, as is released in this paper with `mnms`. Using ACT DR6 data with `mnms`, this paper has constructed empirical noise models and drawn large ensembles of ACT noise simulations. We benchmark `mnms` in appendix I. For future experiments undergoing design and forecast activities, such

as SO, `mnms` can also serve as an emulator of otherwise expensive TOD simulations. Thus, `mnms` could relieve a bottleneck in the simulation-to-forecast feedback loop, enabling iterative optimization of experiment design.

A   Example noise simulations

Figure 23. The difference map (equation (3.6)) for PA5 f090, first split, and simulations from the three noise models (Stokes I).

Example Data and Noise Simulations, Stokes Q

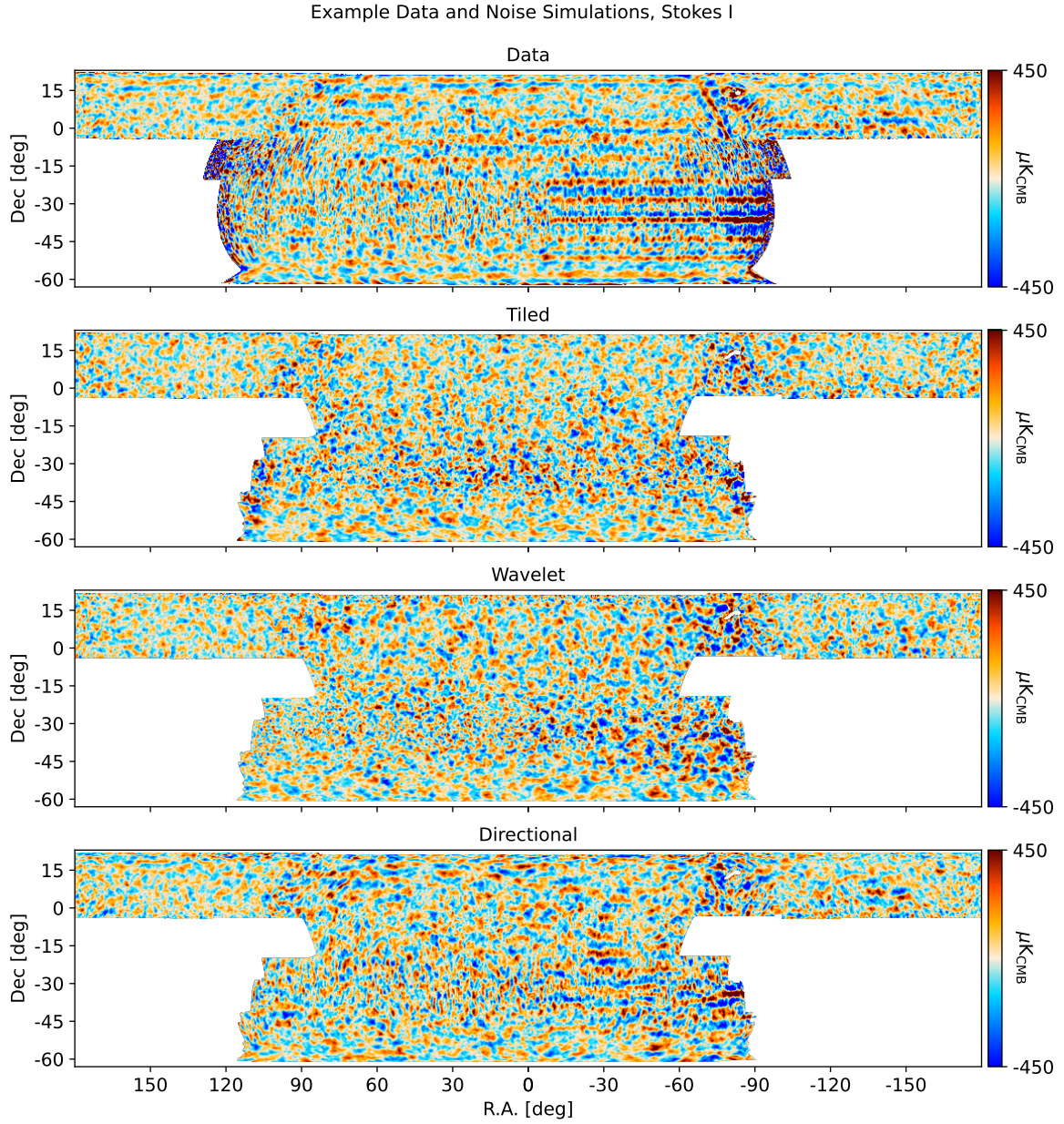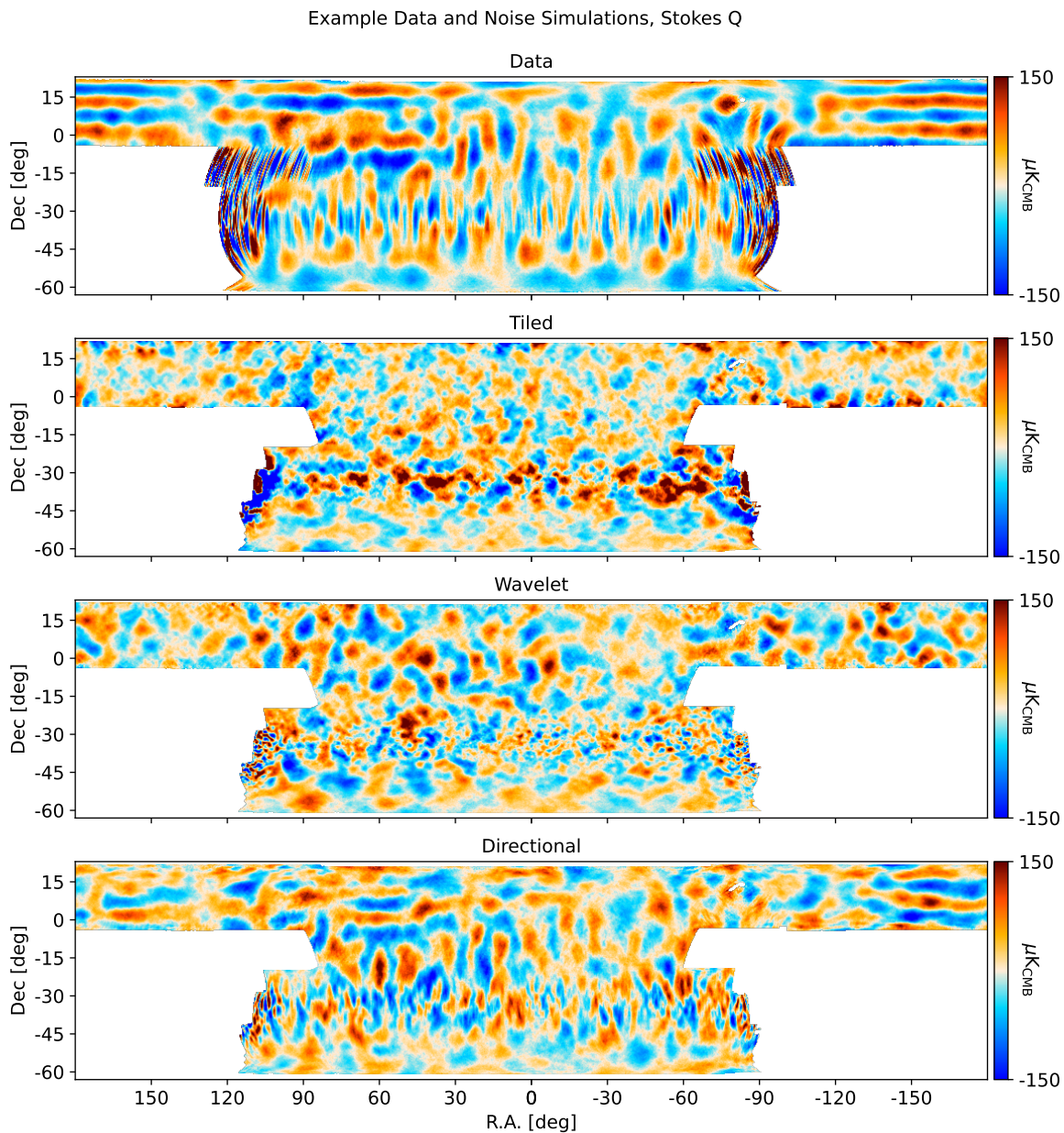**Figure 24**. The difference map (equation (3.6)) for PA5 f090, first split, and simulations from the three noise models (Stokes Q).

B Symbol table

Object Type	Object Shape	Symbol	Definition
Scalar	1	i	Map split index (out of eight)
		a, b	Frequency/polarization index (out of six)
		j	Tile or wavelet index (out of M tiles or wavelets)
Vector	N_{pix}	$\mathbf{f}_{p,i}$	Cross-linking map in split i
		$\boldsymbol{\mu}_{\text{est}}$	Apodized pseudospectrum estimate mask
		$\boldsymbol{\mu}_{\text{bool}}$	Boolean footprint mask
	$6N_{\text{pix}}$	$\mathbf{m}_i$	Observed map in split i, by definition equal to $\mathbf{s} + \mathbf{n}_i$
		$\mathbf{s}$	Signal map, same for all splits i
		$\mathbf{n}_i$	Noise map in split i
		$\mathbf{h}_i$	Inverse-variance map corresponding to $\mathbf{n}_i$
		$\mathbf{c}$	Inverse-variance-weighted coadd map over splits
		$\mathbf{d}_i$	Difference map, $\mathbf{m}_i - \mathbf{c}$ (containing no signal in principle)
		$\mathbf{w}_i$	Inverse-variance map corresponding to $\mathbf{d}_i$
		$\boldsymbol{\nu}_i$	Corrected difference map, $\mathbf{d}_i * \sqrt{\mathbf{w}_i/\mathbf{h}_i}$ (proxy for $\mathbf{n}_i$)
		$\boldsymbol{\xi}_i$	Noise simulation (identically distributed to $\mathbf{n}_i$ in principle)
	$6\sum_{j=0}^{M-1} N_{\text{pix,tile/wav},j}$	$\mathbf{u}_i$	Projection of $\boldsymbol{\nu}_i$ in the tiled or wavelet basis
		$\boldsymbol{\eta}_i$	Gaussian white noise in the tiled or wavelet basis
Matrix	$3N_{\text{pix}} \times 3N_{a_{lm}}$	$\mathbf{Y}$	Spherical harmonic transform (SHT) synthesis
	$N_{\text{pix}} \times N_{\text{pix}}$	$\mathbf{W}$	Quadrature weights for SHT analysis
		$\mathbf{F}$	Discrete Fourier transform (DFT)
	$6N_{a_{lm}} \times 6N_{a_{lm}}$	$\mathbf{C}_{\text{est}}$	Pseudospectra of $\boldsymbol{\mu}_{\text{est}} * \boldsymbol{\nu}_i(*\sqrt{\mathbf{h}_i})$ for the wavelet (tiled) models
	$6N_{\text{pix}} \times 6N_{\text{pix}}$	$\mathbf{H}_i$	Promotion of $\mathbf{h}_i$ to a diagonal matrix
		$\mathbf{N}_i$	Covariance matrix of $\boldsymbol{\xi}_i$ (and $\mathbf{n}_i$ in principle)
	$6N_{\text{pix,tile/wav}} \times 6N_{\text{pix,tile/wav}}$	$\mathcal{N}_{i,j}$	Covariance matrix of the jth tile or wavelet of $\mathbf{u}_i$
	$6\sum_{j=0}^{M-1} N_{\text{pix,tile/wav},j} \times 6N_{\text{pix}}$	$\mathcal{T}_f$	Forward tiling transform, projects a map into tiles or Fourier-space wavelets
	$6\sum_{j=0}^{M-1} N_{a_{lm},\text{wav},j} \times 6N_{a_{lm}}$		Forward tiling transform, projects a map into harmonic-space wavelets

Table 3. Common symbols used throughout sections 3 and 4. Prefactors of 6 in object shapes indicate a vector or matrix for a given detector array with two frequency bands and three Stokes components (I, Q, and U). The prefactor of 3 for the shape of $\mathbf{Y}$ indicates a transformation from T-, E-, and B-mode harmonic coefficients to Stokes I, Q, and U components. As discussed in section 4.1, products of matrices and vectors with incompatible shapes occur according to the `numpy` "broadcasting" rules [42].

C Difference map algebra

We explicitly show the algebra relating to the construction of the difference maps, $\boldsymbol{\nu}_i$. We first prove equation (3.5). Starting from equation (3.4), we have:

$$
\begin{aligned}
\text{Var}(\mathbf{d}_i) = \mathbb{E}(\mathbf{d}_i^2) &= \mathbb{E}\left(\left(\mathbf{n}_i - \frac{\sum_j \mathbf{h}_j * \mathbf{n}_j}{\boldsymbol{\Sigma}}\right)^2\right) = \\
&= \mathbb{E}(\mathbf{n}_i^2) - 2\mathbb{E}\left(\mathbf{n}_i * \frac{\sum_j \mathbf{h}_j * \mathbf{n}_j}{\boldsymbol{\Sigma}}\right) + \mathbb{E}\left(\frac{\sum_i \mathbf{h}_i * \sum_j \mathbf{h}_j * \mathbf{n}_i * \mathbf{n}_j}{\boldsymbol{\Sigma}^2}\right) = \\
&= \mathbb{E}(\mathbf{n}_i^2) - \frac{2\sum_j \mathbf{h}_j * \mathbb{E}(\mathbf{n}_i * \mathbf{n}_j)}{\boldsymbol{\Sigma}} + \frac{\sum_i \mathbf{h}_i * \sum_j \mathbf{h}_j * \mathbb{E}(\mathbf{n}_i * \mathbf{n}_j)}{\boldsymbol{\Sigma}^2}.
\end{aligned}
\tag{C.1}
$$

Now impose the independence of the noise in different splits and the definition of the inverse-variance in split i, $\mathbb{E}(\mathbf{n}_i * \mathbf{n}_j) = 1/\mathbf{h_i} * \delta_{ij}$. Also substitute the definition $\mathbf{\Sigma} \equiv \sum_i \mathbf{h}_i$:

$$\mathrm{Var}(\mathbf{d}_i) = \frac{1}{\mathbf{h}_i} - \frac{2\sum_j \mathbf{h}_j * 1/\mathbf{h_i} * \delta_{ij}}{\mathbf{\Sigma}} + \frac{\sum_i \mathbf{h}_i * \sum_j \mathbf{h}_j * 1/\mathbf{h_i} * \delta_{ij}}{\mathbf{\Sigma}^2} =$$
$$= \frac{1}{\mathbf{h}_i} - \frac{2}{\mathbf{\Sigma}} + \frac{\sum_i \mathbf{h}_i}{\mathbf{\Sigma}^2} = \tag{C.2}$$
$$= \frac{1}{\mathbf{h}_i} - \frac{1}{\mathbf{\Sigma}}.$$

We then prove equation (3.8). First, we note that because $\boldsymbol{\nu}_i \propto \mathbf{d}_i$, the correlation of $\boldsymbol{\nu}_i$ and $\mathbf{n}_i$ is the same as that of $\mathbf{d}_i$ and $\mathbf{n}_i$. Next, we evaluate the covariance of $\mathbf{d}_i$ and $\mathbf{n}_i$:

$$\mathrm{Cov}(\mathbf{d}_i, \mathbf{n}_i) = \mathbb{E}(\mathbf{d}_i * \mathbf{n}_i) = \mathbb{E}\left(\mathbf{n}_i * \left(\mathbf{n}_i - \frac{\sum_j \mathbf{h}_j * \mathbf{n}_j}{\mathbf{\Sigma}}\right)\right) =$$
$$= \mathbb{E}(\mathbf{n}_i^2) - \frac{\sum_j \mathbf{h}_j * \mathbb{E}(\mathbf{n}_i * \mathbf{n}_j)}{\mathbf{\Sigma}} =$$
$$= \frac{1}{\mathbf{h}_i} - \frac{\sum_j \mathbf{h}_j * 1/\mathbf{h_i} * \delta_{ij}}{\mathbf{\Sigma}} = \tag{C.3}$$
$$= \frac{1}{\mathbf{h}_i} - \frac{1}{\mathbf{\Sigma}} =$$
$$= \mathrm{Var}(\mathbf{d}_i).$$

Now evaluate the correlation coefficient directly:

$$\boldsymbol{\rho}_i = \frac{\mathrm{Cov}(\mathbf{d}_i, \mathbf{n}_i)}{\sqrt{\mathrm{Var}(\mathbf{d}_i) * \mathrm{Var}(\mathbf{n}_i)}} = \frac{\mathrm{Var}(\mathbf{d}_i)}{\sqrt{\mathrm{Var}(\mathbf{d}_i) * \mathrm{Var}(\mathbf{n}_i)}} = \sqrt{\frac{\mathrm{Var}(\mathbf{d}_i)}{\mathrm{Var}(\mathbf{n}_i)}} = \sqrt{\mathbf{h}_i * \left(\frac{1}{\mathbf{h}_i} - \frac{1}{\mathbf{\Sigma}}\right)} = \sqrt{1 - \frac{\mathbf{h}_i}{\mathbf{\Sigma}}}.$$
$$\tag{C.4}$$

Note that all of these equations are simply generalizing the familiar "$N-1$" factors from the unbiased sample variance of a set of equally-weighted observations to the case of unequally-weighted observations.

D Derivation of tiling constraint

We offer a derivation of the backward tiling transform constraint (equation (4.16)) that each noise model must satisfy. We will show that it emerges from a desire to not bias the diagonal entries of the full noise covariance matrix. We begin by considering a simpler version of the noise simulation equations (equations (4.18), (4.31), (4.41)). Specifically, we ignore any "exterior" filtering operations, and we replace any DFTs or SHT matrices with a "universal" transform matrix $\mathbf{U}$ (we also drop the split index i for brevity). In this way, a noise simulation from all three noise models appears identical:

$$\boldsymbol{\xi}_{\mathrm{oper}} = \mathcal{T}_b^\dagger \mathbf{U} \mathcal{N}_{\mathrm{sparse}}^{\frac{1}{2}} \boldsymbol{\eta} = \sum_{j=0}^{M-1} \mathbf{P}_j^\dagger \boldsymbol{\tau}_{j,\mathrm{b}}^\dagger \mathbf{U} \mathcal{N}_{j,\mathrm{sparse}}^{\frac{1}{2}} \boldsymbol{\eta}_j. \tag{D.1}$$

where we have generalized the bases of these objects: the simulation is in the "operational" basis where we define and apply the tiling transform profiles — for the tiled/wavelet/directional model this is map space/harmonic space/Fourier space — and the model covariance is in the

"sparse" basis. The matrix $\mathbf{U}$ thus projects a vector from the "sparse" to the "operational" basis — it is a DFT or a SHT matrix.

We want to determine the conditions under which the operational space noise covariance matrix, $\mathbf{N}_{\mathrm{oper}} = \langle \boldsymbol{\xi}_{\mathrm{oper}} \boldsymbol{\xi}_{\mathrm{oper}}^\dagger \rangle$, is unbiased:

$$
\begin{aligned}
\mathbf{N}_{\mathrm{oper}} = \langle \boldsymbol{\xi}_{\mathrm{oper}} \boldsymbol{\xi}_{\mathrm{oper}}^\dagger \rangle &= \sum_{j=0}^{M-1} \sum_{k=0}^{M-1} \mathbf{P}_j^\dagger \boldsymbol{\tau}_{j,\mathrm{b}}^\dagger \mathbf{U} \mathcal{N}_{j,\mathrm{sparse}}^{\frac{1}{2}} \langle \boldsymbol{\eta}_j \boldsymbol{\eta}_k^\dagger \rangle \mathcal{N}_{k,\mathrm{sparse}}^{\frac{1}{2},\dagger} \mathbf{U}^\dagger \boldsymbol{\tau}_{k,\mathrm{b}} \mathbf{P}_k \\
&= \sum_{j=0}^{M-1} \sum_{k=0}^{M-1} \mathbf{P}_j^\dagger \boldsymbol{\tau}_{j,\mathrm{b}}^\dagger \mathbf{U} \mathcal{N}_{j,\mathrm{sparse}}^{\frac{1}{2}} \mathbf{I} \delta_{j,k} \mathcal{N}_{k,\mathrm{sparse}}^{\frac{1}{2},\dagger} \mathbf{U}^\dagger \boldsymbol{\tau}_{k,\mathrm{b}} \mathbf{P}_k \\
&= \sum_{j=0}^{M-1} \mathbf{P}_j^\dagger \boldsymbol{\tau}_{j,\mathrm{b}}^\dagger \mathbf{U} \mathcal{N}_{j,\mathrm{sparse}} \mathbf{U}^\dagger \boldsymbol{\tau}_{j,\mathrm{b}} \mathbf{P}_j \\
&= \sum_{j=0}^{M-1} \mathbf{P}_j^\dagger \boldsymbol{\tau}_{j,\mathrm{b}}^\dagger \mathbf{N}_{j,\mathrm{oper}} \boldsymbol{\tau}_{j,\mathrm{b}} \mathbf{P}_j
\end{aligned}
\tag{D.2}
$$

where we have identified $\mathbf{U} \mathcal{N}_{j,\mathrm{sparse}} \mathbf{U}^\dagger$ with $\mathbf{N}_{j,\mathrm{oper}}$. Now consider a single matrix element indexed by x, x' — that is, the covariance between "pixels" (or "modes") in the operational basis, $N_{\mathrm{oper}}^{x,x'}$. Such an element is formed by taking the xth row of $\mathbf{P}_j^\dagger \boldsymbol{\tau}_{j,\mathrm{b}}^\dagger$ and x'^{th} column of $\boldsymbol{\tau}_{j,\mathrm{b}} \mathbf{P}_j$, each of which have at most one nonzero element. Thus, for a given j, $N_{\mathrm{oper}}^{x,x'}$ receives a contribution from at most one element of $\mathbf{N}_{j,\mathrm{oper}}$; let us index that element by e and e':

$$
\begin{aligned}
N_{\mathrm{oper}}^{x,x'} &= \sum_{j=0}^{M-1} (\mathbf{P}_j^\dagger \boldsymbol{\tau}_{j,\mathrm{b}}^\dagger)_x (\boldsymbol{\tau}_{j,\mathrm{b}} \mathbf{P}_j)_{x'} N_{j,\mathrm{oper}}^{e,e'} \\
N_{\mathrm{oper}}^{x,x'} &= N_{\mathrm{oper}}^{e,e'} \sum_{j=0}^{M-1} (\mathbf{P}_j^\dagger \boldsymbol{\tau}_{j,\mathrm{b}}^\dagger)_x (\boldsymbol{\tau}_{j,\mathrm{b}} \mathbf{P}_j)_{x'} \\
1 &= \sum_{j=0}^{M-1} (\mathbf{P}_j^\dagger \boldsymbol{\tau}_{j,\mathrm{b}}^\dagger)_x (\boldsymbol{\tau}_{j,\mathrm{b}} \mathbf{P}_j)_{x'}.
\end{aligned}
\tag{D.3}
$$

In the case of $x = x'$ (i.e., the diagonal of the map-based covariance matrix), equation (D.3) recovers equation (4.16). We had to use two sleights-of-hand in the second and third lines of equation (D.3). In the second line, we moved $N_{j,\mathrm{oper}}^{e,e'}$ out of the sum, where it became $N_{\mathrm{oper}}^{e,e'}$. In the third line, we asserted that $N_{j,\mathrm{oper}}^{x,x'} = N_{j,\mathrm{oper}}^{e,e'}$. To do so, we assume that any given element of the covariance matrix on the left-hand-side is *equal to* each of its occurrences in each tile or wavelet, j. In reality, each tile or wavelet covariance will disagree on the exact value of a given shared element, but if we assume that these separate estimates are unbiased, then satisfying equation (D.3) will likewise yield an unbiased full covariance matrix.

There are a few more caveats of note. Working backwards, it is difficult to construct useful tiles or wavelets that satisfy equation (D.3) for $x \neq x'$, and indeed for the kernels in this paper this sum approaches zero as the difference between x and x' increases. The upshot is that we cannot model correlations at scales larger than the tile or wavelet size within the tiling scheme itself. Such "long-distance" correlations instead are modeled by the filtering steps that we excluded from the preceding argument. Likewise, we assume that our sparse-basis covariance blocks $\mathcal{N}_{j,\mathrm{sparse}}$ are unbiased estimates, but that is likely not true. For instance, equation (4.9) is the zeroth-order correction to the mode-coupling induced by

the forward tiling transform. Lastly, we must assume that the sparse-basis covariance blocks are themselves uncorrelated, but we know that is not true, since the forward tiling kernels from which the blocks are measured also overlap. In sum, while satisfying equation (4.16) is necessary for a performant noise model, it alone is not sufficient.

E Directional wavelet kernels

As discussed in section 4.4, the directional model utilizes radially- and azimuthally-separable wavelets in the Fourier plane:

$$\tau_{j,\mathrm{f}}(\mathbf{k}) \to \tau_{u,v,\mathrm{f}}(\mathbf{k}) = \kappa_u(k)\eta_{u,v}(\phi) \tag{E.1}$$

where $\mathbf{k}$ is the 2D Fourier mode, $k = |\mathbf{k}|$, $\phi = \arg(\mathbf{k})$, κ are the wavelet kernels of [52, 85] with the parameters given in section 4.3.1, η are the azimuthal kernels, and we have expanded the wavelet index j into a "radial" index u and "azimuthal" index v. Evaluated at each discrete $\mathbf{k}$ in the map Fourier transform, $\tau_{j,\mathrm{f}}$ becomes the diagonal of the forward wavelet kernel matrix $\boldsymbol{\tau}_{j,\mathrm{f}}$. The fact that η also carries the "radial" index denotes that each radial kernel has its own distinct set of associated azimuthal kernels.

We define the azimuthal kernels as follows:

$$\eta_{u,v}(\phi) \to \eta_v(\phi; n_u, p_u) = \sum_{k=-\infty}^{\infty} \begin{cases} A_{k,p}\cos(\omega_{n,p}(\phi - \phi_{0,k,n,v}))^p, & \omega_{n,p}|\phi - \phi_{0,k,n,v}| < \pi/2 \\ 0, & \omega_{n,p}|\phi - \phi_{0,k,n,v}| \geq \pi/2 \end{cases} \tag{E.2}$$

where:

$$\begin{aligned} A_{k,p} &= i^p(-1)^{kp}\sqrt{\frac{2^{2p}}{(p+1)\binom{2p}{p}}} \\ \omega_{n,p} &= \frac{n+1}{p+1} \\ \phi_{0,k,n,v} &= \left(k + \frac{v}{n+1}\right)\pi \end{aligned} \tag{E.3}$$

with $n \geq 0$, $0 \leq p \leq n$, $0 \leq v \leq n$, and all indices are integers. One should interpret this as follows. For each radial index u, we choose two parameters for the set of associated azimuthal kernels: n and p. n sets the number $(n+1)$ of azimuthal kernels, and p is the kernel "shape" parameter. v is the kernel index, and it generates $n+1$ copies of the kernel equally spaced in azimuth (at each $\phi_{0,k,n,v}$). The kernel shape is the p-th power of a cosine centered on $\phi_{0,k,n,v}$, and truncated at either side of $\phi_{0,k,n,v}$ where the $\cos(\ldots)^p$ function intersects the x-axis. The width of the cosine is set by $\omega_{n,p}$. The normalization ensures the kernels satisfy equation (4.16) (with $\tau_{j,\mathrm{b}} = \tau_{j,\mathrm{f}}*$) and respect reality conditions when applied to the Fourier-transformed map in section 4.4.1, step 3. Finally, the sum over k ensures that this piecewise function is periodic in π. Example azimuthal kernels are shown in the top row of figure 25.

The guiding principle in the construction of $\eta_{u,v}$ is the tradeoff in locality between map space and Fourier space: a kernel compact in map space will necessarily be extended in Fourier space, and vice versa. Taking the radial kernels as a given (see figure 12), all else being equal the large-scale wavelets will be therefore be extended in map space, while the small scale wavelets will be compact in map space. We counteract this by tuning the n and p parameters for each radial kernel. For a given number of azimuthal kernels $n+1$, increasing
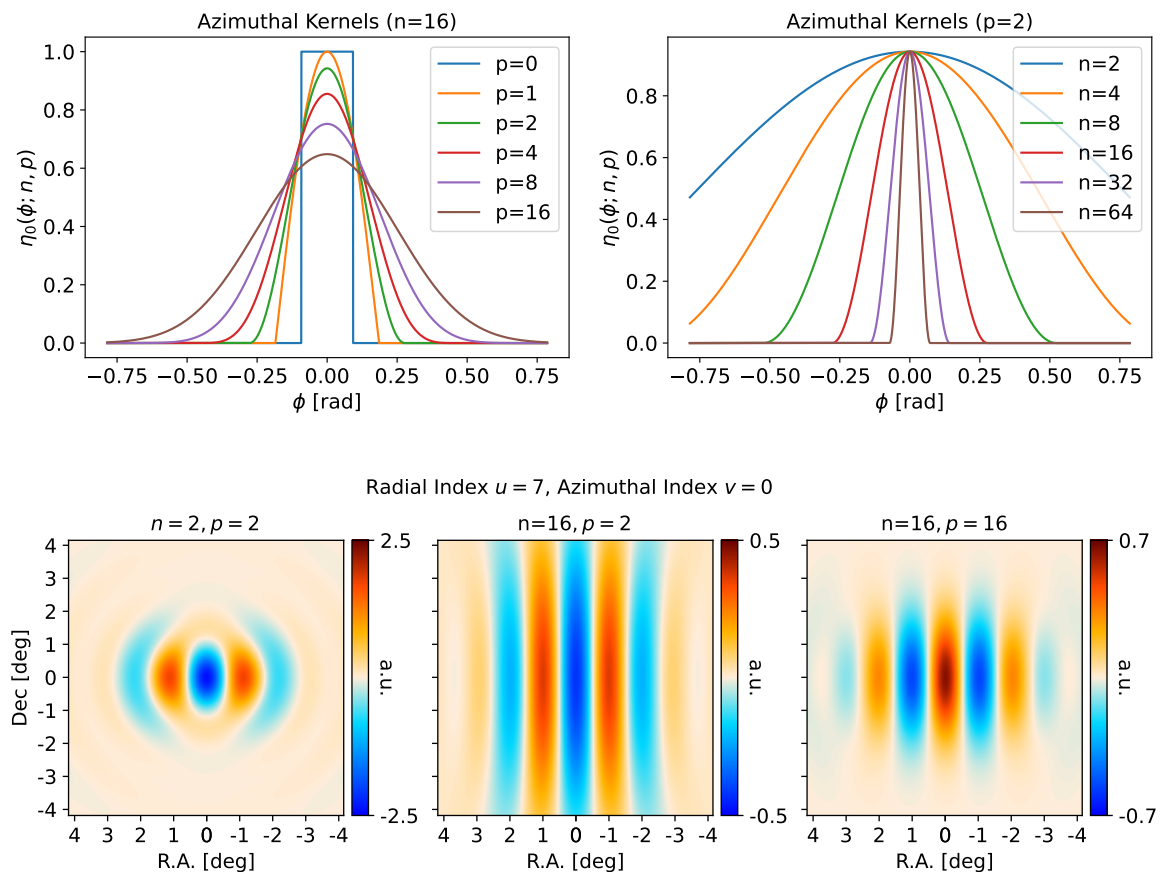
Figure 25. *Top*: 1D profiles of various directional wavelet azimuthal kernels. The left plot demonstrates the decreasing Fourier space compactness with increasing shape parameter, p, values. The right plot demonstrates increasing Fourier space compactness with increasing number of kernels, $n+1$, values. *Bottom*: map space 2D kernels corresponding to $u = 7, v = 0$ for the kernel set used in this paper. The left map space kernel corresponds to the $n = 2, p = 2$ case, which has the most spatial locality and least directional locality. The middle map space kernel corresponds to the $n = 16, p = 2$ case, which has the least spatial locality and most directional locality. The right map space kernel corresponds to the $n = 16, p = 16$ case, which has intermediate spatial and directional locality.

the shape parameter p decreases the compactness of the kernel in Fourier space (and increases its compactness in map space). We note that $p = 0$ corresponds to $n + 1$ equally spaced, non-overlapping tophat functions, while $p = n$ corresponds to the steerable basis of [36]. On the other hand, for a given p, increasing n increases the compactness of the kernels in Fourier space (and decreases its compactness in map space). This tradeoff is evident in the bottom row of figure 25. Therefore, we select low n and high p values for the large-scale wavelets, and the opposite for the small scale wavelets, subject to $0 \leq p \leq n$. Meanwhile, computational interests favor fewer, more-compact Fourier space kernels at high u.

We found the parameters in table 4, which we used in section 4.4, to be a good balance of all considerations. There are fewer radial kernels (16 vs. 26) compared to the isotropic wavelet model to make up for the increased total kernel count. The choice of n and p recalls the "parabolic scaling" in [16], such that the 2D kernels become increasingly directional (narrow) at larger u (see figure 13), except at the highest radial kernel. This allows the

Rad. Kern.	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$\ell_{\min}$	0	7	11	17	27	43	69	110	176	282	451	721	1,153	1,845	2,952	4,723
$\ell_{\max}$	11	17	27	43	69	110	176	282	451	721	1,153	1,845	2,952	4,723	7,556	∞
n	0	6	6	6	6	12	12	12	12	24	24	36	36	36	36	0
p	0	6	4	2	2	12	8	4	2	12	8	2	2	2	2	0

Table 4. Parameters of the directional wavelet kernel set used in this paper. The radial kernels (u-indexed) and associated $\ell_{\min}$, $\ell_{\max}$ values are generated by supplying the parameters $\lambda = 1.6$, $\ell_{\min} = 10$, and $\ell_{\max} = 5,300$ to the wavelet functions prescribed in [52, 85]. We prescribe the n and p values to generate the azimuthal kernel set for each radial kernel.

resolution of sufficient detail in 2D Fourier space (e.g., figure 18) without sacrificing large-scale detail in map space (e.g., figure 20). We note that we accomplished this by increasing n in large steps, while decreasing p within each step. A similar narrowing of the 2D kernels with increasing radii could have been accomplished, for example, by more smoothly increasing n while holding p constant.

F Boolean mask construction

As discussed in section 4.5.1, we exclude some regions of the ACT footprint from our noise models and simulations. We do this through the application of a boolean mask, $\boldsymbol{\mu}_{\rm bool}$. We note that $\boldsymbol{\mu}_{\rm bool}$ can only have a value of 0 or 1 in a given pixel (corresponding to omitting the pixel, and permitting the pixel, respectively). Here, we motivate and define this boolean mask.

The most natural boolean mask is one that selects exactly the pixels that have non-zero hits; however, we construct a slightly more restrictive mask. In addition to unobserved pixels, we mask regions with low cross-linking that abut otherwise well cross-linked/deep regions. These regions are predominantly located on the left and right sides of the southern half of the ACT footprint (the solid black areas figure 15). We found that masking these regions significantly improved the low-ℓ power spectra ratios with respect to the data (see section 5). This can be understood as a fundamental limitation of our modeling: as with residual point sources, our models cannot resolve the step-function-like change from high cross-linking to approximately no cross-linking at the edges of these masked regions. In other words, if we did not mask these shallow, poorly cross-linked regions, their noise power would mix into adjacent deep, well cross-linked regions in the noise models. Importantly, these deep, well cross-linked regions are exactly those retained in the DR6 CMB power spectrum and CMB lensing pipelines — the main consumers of our noise simulations. Therefore, we try to optimize noise model performance in these priority regions, while sacrificing as little data as possible.

We construct $\boldsymbol{\mu}_{\rm bool}$ separately in two disjoint regions. The first (northern) region includes all pixels north of $-6°$ Dec between $180°$ and $125°$ R.A., north of $3°$ Dec between $125°$ and $-100°$ R.A., and again north of $-6°$ Dec between and $-100°$ and $-180°$ R.A.. The second (southern) region is the complement of the northern region. First, in both regions, we set $\boldsymbol{\mu}_{\rm bool} = 1$ in pixels observed by *all* eight splits in *both* frequencies of a given modeled detector array, and 0 otherwise. Then, only in the southern region, we also set $\boldsymbol{\mu}_{\rm bool} = 0$ in pixels whose cross-linking statistic is less than 0.001 in the point-source subtracted coadd map of *any* DR6 array or frequency. A considerable fraction (0.65%) of $\boldsymbol{\mu}_{\rm est}$ in the first region would be cut by the cross-linking threshold, hence why we only apply the cross-linking

cut to the second region. In the second region, the fraction of $\boldsymbol{\mu}_{\text{est}}$ cut by the cross-linking threshold is only 0.04%. We add these pixels in the second region back to $\boldsymbol{\mu}_{\text{bool}}$ such that $\boldsymbol{\mu}_{\text{bool}}$ fully encompasses $\boldsymbol{\mu}_{\text{est}}$. Together, this procedure is effective at removing the large areas of approximately zero cross-linking on the left and right sides of the southern region, while preserving the data used in downstream analyses. For PA5, for instance, this results in the blue outline in figure 15. For the other arrays, the mask is visually indistinguishable.

The somewhat ad-hoc procedure of this section is only necessary given the properties of the ACT scanning strategy cross-linking. Future surveys can mitigate the need for a boolean mask by ensuring more uniform cross-linking.

G   Model smoothing

This appendix expands on our discussion of how and why we smooth our wavelet noise models. As discussed in the text, we use a white-noise map space covariance, modulated by either the mapmaker cross-linking and/or inverse-variance maps, as the fiducial input to our optimization procedure. This simplification has a twofold advantage. First, we showed in section 3.2 that the cross-linking and inverse-variance maps approximately capture the spatial noise structure at large and small scales, respectively. Since we do not know the true noise covariance of the DR6 noise, these products serve as realistic proxies. In particular, by downgrading these maps to specific, lower resolutions (and Nyquist bandlimits), we can emulate the noise covariance at specific wavelet scales. Second, by assuming the noise is white, we can evaluate the agreement between noise models and the input covariance using an analytic likelihood.

We proceed as follows:

1. For a given array and frequency, select either the raw inverse-variance map alone, $\mathbf{h}$, or the combination of $\mathbf{h}$ and the raw cross-linking map, $\mathbf{f}_p$. Using Fourier space resampling, downgrade all maps by a given factor d, such that the Nyquist bandlimit of the maps is $\ell_{\max} = 21,600/d$. In both cases, invert the downgraded inverse-variance map to obtain a "variance" map.

2. In the first case, let the variance map define an underlying, true, white-noise covariance map, $\boldsymbol{\sigma}_0^2 = 1/\mathbf{h}$. This map only has high variance when the inverse-variance is small. In the second case, let the underlying, true, white-noise covariance map be defined as $\boldsymbol{\sigma}_0^2 = \mathbf{f}_p/\mathbf{h}$. Relative to the inverse-variance map, this map has high power when the cross-linking is also poor ($\mathbf{f}_p \sim 1$), following the discussion of equation (3.10).

3. Mask $\boldsymbol{\sigma}_0^2$ with the same boolean mask for the selected array as in section 4.5.3.

4. Draw an uncorrelated, Gaussian realization from $\mathcal{N}(0, \boldsymbol{\sigma}_0^2)$, called $\mathbf{s}_i$. Let this be analogous to the hypothetical i-th independent realization of the noise in the data.

5. Feed $\mathbf{s}_i$ into one of several noise modeling routines, utilizing a fixed FWHM smoothing factor f in all models:

   - A white-noise model which is simply the pixel-wise square of $\mathbf{s}_i$, $\mathbf{s}_{i,\text{model}}^2$.
   - The isotropic wavelet model of section 4.3, without the pseudospectrum-filtering step.
   - The directional wavelet model of section 4.4, without the pseudospectrum-filtering step.
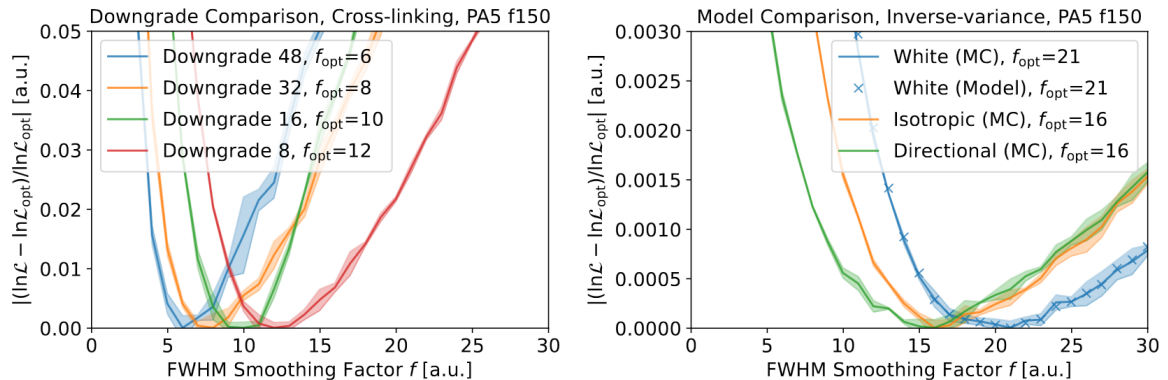
Figure 26. Model likelihoods $(-\ln\mathcal{L})$ as a function of smoothing factor f. Solid lines (shaded bands) denote the mean (1σ range) over the $N_i = 3$ noise realizations from the fiducial input covariance. Curves have been recentered and normalized relative to their minima to facilitate direct comparison. *Left*: likelihoods using the PA5 f150 cross-linking and inverse-variance combination map $(\boldsymbol{\sigma}_0^2 = \mathbf{f}_p/\mathbf{h})$ as the input covariance. Different colors denote different choices of downgrade factors. *Right*: likelihoods for the set of noise modeling routines enumerated in the appendix, using the PA5 f150 inverse-variance map $(\boldsymbol{\sigma}_0^2 = 1/\mathbf{h})$ downgraded by a factor of four as the input covariance. The data labeled "White (MC)" ("White (Model)") refer to the white-noise model's $\mathbf{s}_{i,\mathrm{MC}}^2$ map ($\mathbf{s}_{i,\mathrm{model}}^2$ map). The agreement between the curve and overplotted crosses demonstrates that 30 simulations are sufficient to reach convergence of Monte Carlo estimates. For the other models, we only have the option of using the Monte Carlo estimates. For both panels, the relatively small scatter over the $N_i = 3$ noise realizations indicates the optima are robust.

6. Draw $N_{\mathrm{sim}} = 30$ simulations, $\mathbf{s}_{i,\mathrm{sim}}$, and form the Monte Carlo estimated, white-noise covariance map $\mathbf{s}_{i,\mathrm{MC}}^2 = 1/N_{\mathrm{sim}} \sum \mathbf{s}_{i,\mathrm{sim}}^2$. For the two wavelet models, this is necessary because it is not possible to directly compare the model itself, which is in the wavelet basis, to the input covariance, $\boldsymbol{\sigma}_0^2$, which is in the map basis. For the white-noise model, as noted above, we retain both the Monte Carlo estimate, $\mathbf{s}_{i,\mathrm{MC}}^2$, as well the model itself, $\mathbf{s}_{i,\mathrm{model}}^2$, since the model is in the map basis. By comparing results using $\mathbf{s}_{i,\mathrm{MC}}^2$ to $\mathbf{s}_{i,\mathrm{model}}^2$ in the next step, this allows us to check that $N_{\mathrm{sim}} = 30$ simulations is sufficient for the Monte Carlo estimates to converge (e.g., note the agreement between the blue crosses to the blue line in figure 26).

7. For the white-noise model, the isotropic wavelet model, and the directional wavelet model, evaluate the likelihood of $\mathbf{s}_{i,\mathrm{MC}}^2$ given the true covariance $\boldsymbol{\sigma}_0^2$, $\mathcal{L}(\mathbf{s}_{i,\mathrm{MC}}^2|\boldsymbol{\sigma}_0^2)$. For the white-noise model only, also evaluate $\mathcal{L}(\mathbf{s}_{i,\mathrm{model}}^2|\boldsymbol{\sigma}_0^2)$. To avoid the poorly-observed edges of $\boldsymbol{\sigma}_0^2$, only perform the evaluation using pixels where the pseudospectrum mask from section 4.5.3 is nonzero.

8. Accumulate results over $N_i = 3$ input noise realizations.

The likelihood, $\mathcal{L}(\mathbf{s}_{i,\mathrm{MC}}^2|\boldsymbol{\sigma}_0^2)$, follows the gamma distribution in each pixel x:

$$-\ln\mathcal{L}(\mathbf{s}_i^2|\boldsymbol{\sigma}_0^2) = -\sum_x \mathcal{P}_{\mathrm{gamma}}(s_i^2|\alpha = N_{\mathrm{sim}}/2, \theta = 2\sigma_{0,x}^2/N_{\mathrm{sim}}) \tag{G.1}$$

where $\mathcal{P}_{\mathrm{gamma}}$ is the gamma probability density with shape parameter α and scale parameter θ. That is, for a given pixel with variance σ_0^2, we evaluate the likelihood that the average

of $N_{\text{sim}} = 30$ independent samples of that variance equals the observed value. We minimize $-\ln\mathcal{L}$ as a function of the smoothing factor f, for various modeling routines and downgrade factors d. Example likelihoods are shown in the left panel of figure 26.

As a motivating example, we dispel the misconception that noise models with more wavelet kernels, and thus with more parameters, require increased model smoothing. This view arises from the observation that, in the case of a single-frequency, unpolarized map, the directional model, with its 280 wavelet kernels, contains a factor $\sim$53 more parameters than the input data dimension. Compare that to the white-noise model we enumerated above, which has exactly the same number of parameters as the input data dimension. However, this argument overlooks that these 280 wavelet kernels are compact in Fourier space: their corresponding wavelet maps will therefore be smooth, with highly-correlated parameters. In fact, they will be correlated by exactly the right amount to guarantee that the number of degrees-of-freedom will be equal to the input data dimension. This is a consequence of equation (4.16): with more kernels supporting a single element of the map space covariance matrix, equation (4.16) averages over equally more terms, and with proportionally smaller weights.

To confirm this expectation, we perform the above procedure for each of the listed noise modeling routines. We use inverse-variance maps downgraded by a factor of four as inputs, and examine the results in the right panel of figure 26. On the whole, we find the likelihood is generally insensitive to the noise modeling routine. In fact, the likelihood slightly prefers less smoothing as the number of wavelet kernels increases. This preference can also be understood as a consequence of the spatially-smooth wavelet kernels: they have a fundamentally limited ability to localize features in map space, and so enter the regime of "over-smoothing" sooner than a perfectly-local white-noise model. We factor this limitation into our final choice of smoothing factors in section 4.5.4.

H Benefits of pseudospectrum filter: mode-decoupling and stationarity

The noise pseudospectrum filter is necessary for the construction of minimally biased noise models. We demonstrate that its exclusion leads to significant deviations from nominal in ratios of simulated versus input power spectra. We also demonstrate that our model estimation and simulation algorithms — when including the filter — suppress these deviations even if the filter is suboptimal. We examine the performance of a tiled and wavelet model in both cases.

We proceed as follows:

1. We draw a full-sky, isotropic realization from an input fiducial noise power spectrum (bandlimited to $\ell_{\text{max}} = 5,400$).

   - For the no-filter, tiled (wavelet) model case, we use the PA5 f090 noise EE (TT) pseudospectrum.

   - For the supoptimal-filter case, both models, we construct the input spectrum from scratch. The spectrum is much closer to flat than the actual noise spectra, but is not exactly flat (see figure 27). It is equivalent to the situation where we have drawn from the same unfiltered PA5 f090 noise spectra, and then applied an imperfect noise pseudospectrum filter, leaving some residual ℓ-dependence to the filtered noise realization.
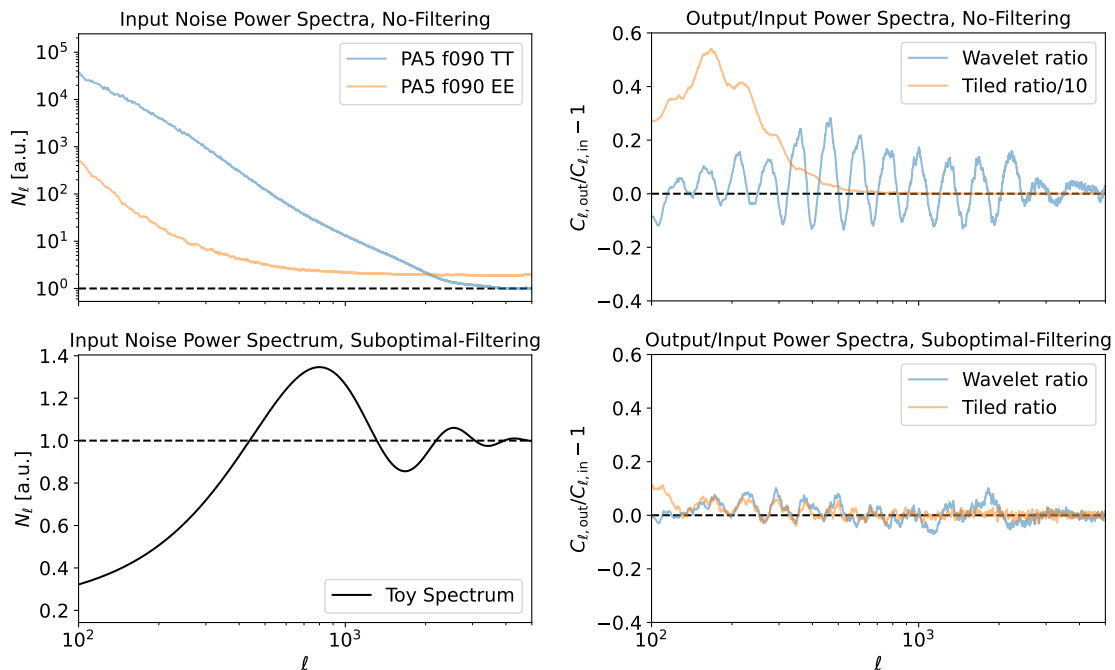
Figure 27. *Top-left*: noise pseudospectra from the actual data used as input power spectra for the no-filtering case. They are the mean over splits from the same curves as figure 3. The black dashed line indicates an arbitrary normalization to the spectra which has no effect on the results. *Top-right*: power spectra ratios between tiled and wavelet simulations, and their input noise realizations. The input realization from the tiled (wavelet) model is drawn from the EE (TT) spectrum. The tiled model deviation has been divided by 10 to fit on the same plot. Power spectra entering the ratios are smoothed with a tophat kernel of size $\Delta\ell = 25$. *Bottom-left*: toy spectrum used as input for the suboptimal-filtering case. Now the black dashed line signifies that this spectrum represents the effect of an imperfect filter. *Bottom-right*: power spectra ratios for tiled and wavelet simulations, built off the same toy realization. The tiled model deviation is no longer divided by 10. Again, power spectra entering the ratios are smoothed with a tophat kernel of size $\Delta\ell = 25$.

2. We mask this realization with the same boolean footprint mask as the actual PA5 f090 data (see section 4.5.3).

3. From these realizations we generate a noise model as in section 4, with some simplifications:

   - For the tiled model, we exclude the inverse-variance map filter step since the input noise realization is isotropic.
   - For all cases, we exclude the pseudospectrum filter step.
   - For all cases, we do not smooth the models at all. This is especially important for the tiled model whose sparse basis is in Fourier space: smoothing directly mixes power across scales and could also produce deviations in output/input power spectra ratios. For consistency, we do not smooth the wavelet model.

4. From these models, we draw a single simulation as in section 4, with the analogous simplifications.

5. We compare the pseudospectra of the ouput noise simulation with the input noise realization, as measured in the apodized pseudospectrum mask of section 4.5.3.

The downsides of ignoring the pseudospectrum filter are self-evident in figure 27. As discussed in section 4.2.1, multiplying each tile by an apodized mask in map space mixes modes in Fourier space. When the power spectrum of the signal in the tile is especially steep, as in the case of the PA5 f090 EE noise, a given mode receives excess power from nearby modes at slightly larger scales. In the simple demonstration here, we observe an excess in simulation power of $\sim 500\%$ near $\ell = 200$ relative to the input noise power. Meanwhile, as discussed in section 4.3.1, the wavelet model assumes stationary noise in harmonic space; this assumption is clearly violated by the strong ℓ-dependence of the PA5 f090 TT noise power spectrum. Instead, the model projects out the stationary part of the noise — that is, without an absolute ℓ-dependence — within each wavelet kernel. This produces a "staircase-like" noise model in harmonic space, with a flat step on each wavelet kernel, rather than the smoothly decreasing input spectrum. Thus, we observe an oscillating ratio between the simulation and the input realization

By including a (sub)optimal pseudospectrum filter we suppress these issues. As figure 27 shows, even if the power spectrum of the input noise contains $\mathcal{O}(1)$ variations over ℓ, simulations drawn from the noise models are less biased over a large range of angular scales. Such a power spectrum could plausibly arise in section 4 after the pseudospectrum filtering step in regions of sky whose local noise power spectrum does not match the one used in the filter. We judge these improvements to be worth the cost of the filter's limited map space resolution discussed in section 6. Finally, we note that when applied to the real data, the pseudospectrum filter does not guarantee unbiased simulated power spectra, even if the filter perfectly matches the actual noise power spectrum. Rather, this appendix goes to show that such a filter is necessary, but not sufficient, for a performant noise model.

I   Code performance table

Model Type	Model Estimation				Simulation			
	Memory (GB)	Disk (GB)	Wall Time (s)		Memory (GB)	Disk (MB)	Wall Time (s)	
			10 CPU threads	40 CPU threads			10 CPU threads	40 CPU threads
Tiled	33.0	5.5	236	162	15.8	638	11	6
Wavelet	28.7	8.0	371	237	24.1	638	28	18
Directional	193.8	149.9	1601	750	185.9	638	76	42

Table 5. `mnms` computational benchmarks for noise models and simulations with a bandlimit of $\ell_{\max} = 5,400$ as in the rest of the paper. All jobs were run on a single Intel Xeon Gold 6242R CPU node with 40 available threads and 380 GB of available memory. "Model estimation" refers to the construction of the sparse covariance matrix in sections 4.2.1, 4.3.1, and 4.4.1. We ran one such job each utilizing 10 or 40 threads. "Simulation" refers to the drawing of a noise simulation in sections 4.2.2, 4.3.2, and 4.4.2. We ran one job each utilizing 10 or 40 threads and generating four simulations. "Memory" is the total memory utilized during a job writing the model or simulations to disk, including I/O and preprocessing (averaged over the two jobs). "Disk" is the size of the written model or one simulation. "Wall Time" is the clock time spent constructing the model or drawing one simulation (the median time of the four simulations), excluding I/O and preprocessing. The directional model's memory and disk requirements are larger than the others' due to the non-compact "cropping" of each wavelet as discussed in section 4.4.1. That is, each wavelet retains the origin of Fourier space, such that (for example) the small-scale wavelets contain nearly all the Fourier modes in the map. Nevertheless, all the benchmarks here (including the directional model) are easily handled by the node and should be subdominant to most downstream analysis pipelines.

References

[1] CMB-S4 collaboration, *CMB-S4 Technology Book, First Edition*, `arXiv:1706.02464` [ɪɴSPIRE].

[2] ACT collaboration, *The Atacama Cosmology Telescope: DR4 Maps and Cosmological Parameters*, *JCAP* **12** (2020) 047 [`arXiv:2007.07288`] [ɪɴSPIRE].

[3] PLANCK collaboration, *Planck intermediate results. LVII. Joint Planck LFI and HFI data processing*, *Astron. Astrophys.* **643** (2020) A42 [`arXiv:2007.04997`] [ɪɴSPIRE].

[4] E. Allys, T. Marchand, J.F. Cardoso, F. Villaescusa-Navarro, S. Ho and S. Mallat, *New Interpretable Statistics for Large Scale Structure Analysis and Generation*, *Phys. Rev. D* **102** (2020) 103506 [`arXiv:2006.06298`] [ɪɴSPIRE].

[5] SPT collaboration, *SPT-3G: A Multichroic Receiver for the South Pole Telescope*, *J. Low Temp. Phys.* **193** (2018) 1057 [ɪɴSPIRE].

[6] ASTROPY collaboration, *Astropy: A Community Python Package for Astronomy*, *Astron. Astrophys.* **558** (2013) A33 [`arXiv:1307.6212`] [ɪɴSPIRE].

[7] ASTROPY collaboration, *The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package*, *Astron. J.* **156** (2018) 123 [`arXiv:1801.02634`] [ɪɴSPIRE].

[8] ASTROPY collaboration, *The Astropy Project: Sustaining and Growing a Community-oriented Open-source Project and the Latest Major Release (v5.0) of the Core Package*, *Astrophys. J.* **935** (2022) 167 [`arXiv:2206.14220`].

[9] D.R. Barron et al., *Conceptual design of the modular detector and readout system for the CMB-S4 survey experiment*, *Proc. SPIE* **12190** (2022) 121900B [`arXiv:2208.02284`] [ɪɴSPIRE].

[10] B. Behnel et al., *Cython: The Best of Both Worlds*, *Comput. Sci. Eng.* **13** (2011) 31.

[11] B.A. Benson et al., *SPT-3G: a next-generation cosmic microwave background polarization experiment on the South Pole telescope*, *Proc. SPIE* **9153** (2014) 552 [`arXiv:1407.2973`] [ɪɴSPIRE].

[12] L.E. Bleem et al., *CMB/kSZ and Compton-y Maps from 2500 deg2 of SPT-SZ and Planck Survey Data*, *Astrophys. J. Suppl.* **258** (2022) 36 [`arXiv:2102.05033`] [ɪɴSPIRE].

[13] M.L. Brown, P.G. Castro and A.N. Taylor, *CMB temperature and polarisation pseudo-C(l) estimators and covariances*, *Mon. Not. Roy. Astron. Soc.* **360** (2005) 1262 [`astro-ph/0410394`] [ɪɴSPIRE].

[14] M. Bucher et al., *Filling in cosmic microwave background map missing data using constrained Gaussian realizations*, *Mon. Not. Roy. Astron. Soc.* **424** (2012) 1694 [`arXiv:1109.0286`].

[15] E. Camphuis, K. Benabed, S. Galli, E. Hivon and M. Lilley, *Accurate cosmic microwave background covariance matrices: Exact calculation and approximations*, *Astron. Astrophys.* **668** (2022) A62 [`arXiv:2204.13721`] [ɪɴSPIRE].

[16] J.Y.H. Chan, B. Leistedt, T.D. Kitching and J.D. McEwen, *Second-Generation Curvelets on the Sphere*, *IEEE Trans. Signal Process.* **65** (2017) 5 [`arXiv:1511.05578`] [ɪɴSPIRE].

[17] S. Cheng, Y.-S. Ting, B. Ménard and J. Bruna, *A new approach to observational cosmology using the scattering transform*, *Mon. Not. Roy. Astron. Soc.* **499** (2020) 5902 [`arXiv:2006.08561`] [ɪɴSPIRE].

[18] S.K. Choi et al., *Characterization of the Mid-Frequency Arrays for Advanced ACTPol*, *J. Low Temp. Phys.* **193** (2018) 267 [`arXiv:1711.04841`] [ɪɴSPIRE].

[19] ACT collaboration, *The Atacama Cosmology Telescope: a measurement of the Cosmic Microwave Background power spectra at 98 and 150 GHz*, *JCAP* **12** (2020) 045 [`arXiv:2007.07289`] [ɪɴSPIRE].

[20] SPT collaboration, *Maps of the Southern Millimeter-wave Sky from Combined 2500 deg^2 SPT-SZ and Planck Temperature Data*, *Astrophys. J. Suppl.* **239** (2018) 10 [arXiv:1803.10682] [INSPIRE].

[21] A. Collette, *Python and HDF5*, O'Reilly Media (2013), https://www.h5py.org/.

[22] F. Couchot et al., *Cosmology with the cosmic microwave background temperature-polarization correlation*, *Astron. Astrophys.* **602** (2017) A41 [arXiv:1609.09730] [INSPIRE].

[23] ACT collaboration, *The Atacama Cosmology Telescope: High-resolution component-separated maps across one-third of the sky*, arXiv:2307.01258 [INSPIRE].

[24] S. Dahal et al., *Four-year Cosmology Large Angular Scale Surveyor (CLASS) Observations: On-sky Receiver Performance at 40, 90, 150, and 220 GHz Frequency Bands*, *Astrophys. J.* **926** (2022) 33 [arXiv:2107.08022] [INSPIRE].

[25] F.A. Dahlen and F.J. Simons, *Spectral estimation on a sphere in geophysics and cosmology*, *Geophys. J. Int.* **174** (2008) 774 [arXiv:0705.3083] [INSPIRE].

[26] R. de Belsunce, S. Gratton, W. Coulton and G. Efstathiou, *Inference of the optical depth to reionization from low multipole temperature and polarization Planck data*, *Mon. Not. Roy. Astron. Soc.* **507** (2021) 1072 [arXiv:2103.14378] [INSPIRE].

[27] J. Delabrouille, J.F. Cardoso, M.L. Jeune, M. Betoule, G. Fay and F. Guilloux, *A full sky, low foreground, high resolution CMB map from WMAP*, *Astron. Astrophys.* **493** (2009) 835 [arXiv:0807.0773] [INSPIRE].

[28] J. Dunkley et al., *Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Bayesian Estimation of CMB Polarization Maps*, *Astrophys. J.* **701** (2009) 1804 [arXiv:0811.4280] [INSPIRE].

[29] J. Dunkley et al., *The Atacama Cosmology Telescope: likelihood for small-scale CMB data*, *JCAP* **07** (2013) 025 [arXiv:1301.0776] [INSPIRE].

[30] R. Dunner et al., *The Atacama Cosmology Telescope: Data Characterization and Map Making*, *Astrophys. J.* **762** (2013) 10 [arXiv:1208.0050] [INSPIRE].

[31] G. Efstathiou, *Hybrid estimation of cmb polarization power spectra*, *Mon. Not. Roy. Astron. Soc.* **370** (2006) 343 [astro-ph/0601107] [INSPIRE].

[32] G. Efstathiou and S. Gratton, *A Detailed Description of the CamSpec Likelihood Pipeline and a Reanalysis of the Planck High Frequency Maps*, arXiv:1910.00483 [INSPIRE].

[33] H.K. Eriksen, J.B. Jewell, C. Dickinson, A.J. Banday, K.M. Gorski and C.R. Lawrence, *Joint Bayesian component separation and CMB power spectrum estimation*, *Astrophys. J.* **676** (2008) 10 [arXiv:0709.1058] [INSPIRE].

[34] POLARBEAR collaboration, *Modeling atmospheric emission for CMB ground-based observations*, *Astrophys. J.* **809** (2015) 63 [arXiv:1501.07911] [INSPIRE].

[35] J.M. Fowler et al., *Optical Design of the Atacama Cosmology Telescope and the Millimeter Bolometric Array Camera*, *Appl. Opt.* **46** (2007) 3444 [astro-ph/0701020] [INSPIRE].

[36] W.T. Freeman and E.H. Adelson, *The design and use of steerable filters*, *IEEE Trans. Pattern Anal. Mach. Intell.* **13** (1991) 891.

[37] N. Galitzki et al., *The Simons Observatory: instrument overview*, *Proc. SPIE* **10708** (2018) 3 [arXiv:1808.04493].

[38] C. García-García, D. Alonso and E. Bellini, *Disconnected pseudo-C$_\ell$ covariances for projected large-scale structure data*, *JCAP* **11** (2019) 043 [arXiv:1906.11765] [INSPIRE].

[39] H. Gomersall, *pyFFTW*, http://dx.doi.org/10.5281/ZENODO.59508 (2016).

[40] K.M. Górski, E. Hivon, A.J. Banday, B.D. Wandelt, F.K. Hansen, M. Reinecke et al., *HEALPix - A Framework for high resolution discretization, and fast analysis of data distributed on the sphere*, *Astrophys. J.* **622** (2005) 759 [astro-ph/0409513] [INSPIRE].

[41] D. Hanson, G. Rocha and K. Gorski, *Lensing reconstruction from PLANCK sky maps: inhomogeneous noise*, *Mon. Not. Roy. Astron. Soc.* **400** (2009) 2169 [arXiv:0907.1927] [INSPIRE].

[42] C.R. Harris et al., *Array programming with NumPy*, *Nature* **585** (2020) 357 [arXiv:2006.10256] [INSPIRE].

[43] S.W. Henderson et al., *Advanced ACTPol Cryogenic Detector Arrays and Readout*, *J. Low Temp. Phys.* **184** (2016) 772 [arXiv:1510.02809] [INSPIRE].

[44] SPT collaboration, *Measurements of the Temperature and E-Mode Polarization of the CMB from 500 Square Degrees of SPTpol Data*, *Astrophys. J.* **852** (2018) 97 [arXiv:1707.09353] [INSPIRE].

[45] E. Hivon, K.M. Gorski, C.B. Netterfield, B.P. Crill, S. Prunet and F. Hansen, *Master of the cosmic microwave background anisotropy power spectrum: a fast method for statistical analysis of large and complex cosmic microwave background data sets*, *Astrophys. J.* **567** (2002) 2 [astro-ph/0105302] [INSPIRE].

[46] S.-P. Patty Ho et al., *Highly uniform 150 mm diameter multichroic polarimeter array deployed for CMB detection*, *Proc. SPIE* **9914** (2017) 991418.

[47] J.D. Hunter, *Matplotlib: A 2D Graphics Environment*, *Comput. Sci. Eng.* **9** (2007) 90 [INSPIRE].

[48] WMAP collaboration, *Three-year Wilkinson Microwave Anisotropy Pr (WMAP) observations: beam profiles, data processing, radiometer characterization and systematic error limits*, *Astrophys. J. Suppl.* **170** (2007) 263 [astro-ph/0603452] [INSPIRE].

[49] J.-H. Kang Ho et al., *2017 upgrade and performance of BICEP3: a 95GHz refracting telescope for degree-scale CMB polarization*, *Proc. SPIE* **10708** (2018) 107082N [arXiv:1808.00567] [INSPIRE].

[50] M.H. Kesden, A. Cooray and M. Kamionkowski, *Lensing reconstruction with CMB temperature and polarization*, *Phys. Rev. D* **67** (2003) 123507 [astro-ph/0302536] [INSPIRE].

[51] S. K. Lam, A. Pitrou and S. Seibert, *Numba: a LLVM-based Python JIT compiler*, in *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC, LLVM '15*, New York, U.S.A. (2015), https://doi.org/10.1145/2833157.2833162.

[52] B. Leistedt, J.D. McEwen, P. Vandergheynst and Y. Wiaux, *S2LET: A code to perform fast wavelet analysis on the sphere*, *Astron. Astrophys.* **558** (2013) A128 [arXiv:1211.1680] [INSPIRE].

[53] Z. Li et al., *The Simons Observatory: a new open-source power spectrum pipeline applied to the Planck legacy data*, arXiv:2112.13839 [INSPIRE].

[54] T. Louis, S. Naess, X. Garrido and A. Challinor, *Fast computation of angular power spectra and covariances of high-resolution cosmic microwave background maps using the Toeplitz approximation*, *Phys. Rev. D* **102** (2020) 123538 [arXiv:2010.14344] [INSPIRE].

[55] ACTPOL collaboration, *The Atacama Cosmology Telescope: Two-Season ACTPol Spectra and Parameters*, *JCAP* **06** (2017) 031 [arXiv:1610.02360] [INSPIRE].

[56] M.S. Madhavacheril, K.M. Smith, B.D. Sherwin and S. Naess, *CMB lensing power spectrum estimation without instrument noise bias*, arXiv:2011.02475 [INSPIRE].

[57] M.S. Madhavacheril et al., *Atacama Cosmology Telescope: Component-separated maps of CMB temperature and the thermal Sunyaev-Zel'dovich effect*, *Phys. Rev. D* **102** (2020) 023534 [arXiv:1911.05717] [InSPIRE].

[58] M. Mallaby-Kay et al., *The Atacama Cosmology Telescope: Summary of DR4 and DR5 Data Products and Data Access*, *Astrophys. J. Suppl.* **255** (2021) 11 [arXiv:2103.03154] [InSPIRE].

[59] J.D. McEwen, C. Durastanti and Y. Wiaux, *Localisation of directional scale-discretised wavelets on the sphere*, *Appl. Comput. Harmon. Anal.* **44** (2018) 59.

[60] M. Millea et al., *Optimal Cosmic Microwave Background Lensing Reconstruction and Parameter Estimation with SPTpol Data*, *Astrophys. J.* **922** (2021) 259 [arXiv:2012.01709] [InSPIRE].

[61] T.W. Morris et al., *The Atacama Cosmology Telescope: Modeling bulk atmospheric motion*, *Phys. Rev. D* **105** (2022) 042004 [arXiv:2111.01319] [InSPIRE].

[62] ACTPol collaboration, *The Atacama Cosmology Telescope: CMB Polarization at $200 < \ell < 9000$*, *JCAP* **10** (2014) 007 [arXiv:1405.5524] [InSPIRE].

[63] S. Naess et al., *The Atacama Cosmology Telescope: arcminute-resolution maps of 18 000 square degrees of the microwave sky from ACT 2008–2018 data combined with Planck*, *JCAP* **12** (2020) 046 [arXiv:2007.07290] [InSPIRE].

[64] Planck collaboration, *Planck 2018 results. I. Overview and the cosmological legacy of Planck*, *Astron. Astrophys.* **641** (2020) A1 [arXiv:1807.06205] [InSPIRE].

[65] Planck collaboration, *Planck 2018 results. III. High Frequency Instrument data processing and frequency maps*, *Astron. Astrophys.* **641** (2020) A3 [arXiv:1807.06207] [InSPIRE].

[66] Planck collaboration, *Planck 2018 results. V. CMB power spectra and likelihoods*, *Astron. Astrophys.* **641** (2020) A5 [arXiv:1907.12875] [InSPIRE].

[67] Planck collaboration, *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6 [arXiv:1807.06209] [*Erratum ibid.* **652** (2021) C4] [InSPIRE].

[68] Planck collaboration, *Planck 2015 results. XI. CMB power spectra, likelihoods, and robustness of parameters*, *Astron. Astrophys.* **594** (2016) A11 [arXiv:1507.02704] [InSPIRE].

[69] Planck collaboration, *Planck 2015 results. XII. Full Focal Plane simulations*, *Astron. Astrophys.* **594** (2016) A12 [arXiv:1509.06348] [InSPIRE].

[70] Planck collaboration, *Planck intermediate results. XLVIII. Disentangling Galactic dust emission and cosmic infrared background anisotropies*, *Astron. Astrophys.* **596** (2016) A109 [arXiv:1605.09387] [InSPIRE].

[71] ACT collaboration, *The Atacama Cosmology Telescope: A Measurement of the DR6 CMB Lensing Power Spectrum and its Implications for Structure Growth*, arXiv:2304.05202 [InSPIRE].

[72] B.R.-S. Blancard, E. Allys, F. Boulanger, F. Levrier and N. Jeffrey, *A new approach for the statistical denoising of Planck interstellar dust polarization data*, *Astron. Astrophys.* **649** (2021) L18 [arXiv:2102.03160] [InSPIRE].

[73] M. Reinecke and D.S. Seljebotn, *Libsharp — spherical harmonic transforms revisited*, *Astron. Astrophys.* **554** (2013) A112.

[74] K.K. Rogers, H.V. Peiris, B. Leistedt, J.D. McEwen and A. Pontzen, *Spin-SILC: CMB polarization component separation with spin wavelets*, *Mon. Not. Roy. Astron. Soc.* **463** (2016) 2310 [arXiv:1605.01417] [InSPIRE].

[75] M. Salatino et al., *The design of the Ali CMB Polarization Telescope receiver*, *Proc. SPIE* **11453** (2020) 114532A [arXiv:2101.09608] [InSPIRE].

[76] A. Schillaci et al., *Design and Performance of the First BICEP Array Receiver*, *J. Low Temp. Phys.* **199** (2020) 976 [`arXiv:2002.05228`] [ɪɴSPIRE].

[77] D.S. Seljebotn, T. Bærland, H.K. Eriksen, K.A. Mardal and I.K. Wehus, *Multi-resolution Bayesian CMB component separation through Wiener-filtering with a pseudo-inverse preconditioner*, *Astron. Astrophys.* **627** (2019) A98 [`arXiv:1710.00621`] [ɪɴSPIRE].

[78] K.M. Smith, O. Zahn and O. Dore, *Detection of Gravitational Lensing in the Cosmic Microwave Background*, *Phys. Rev. D* **76** (2007) 043510 [`arXiv:0705.3980`] [ɪɴSPIRE].

[79] K.M. Smith and M. Zaldarriaga, *Algorithms for bispectra: Forecasting, optimal analysis, and simulation*, *Mon. Not. Roy. Astron. Soc.* **417** (2011) 2 [`astro-ph/0612571`] [ɪɴSPIRE].

[80] POLARBEAR collaboration, *The POLARBEAR-2 and the Simons Array Experiment*, *J. Low Temp. Phys.* **184** (2016) 805 [`arXiv:1512.07299`] [ɪɴSPIRE].

[81] T.L. Svalheim et al., *BEYONDPLANCK — XIV. Polarized foreground emission between 30 and 70 GHz*, *Astron. Astrophys.* **675** (2023) A14 [`arXiv:2011.08503`] [ɪɴSPIRE].

[82] S. Takakura et al., *Performance of a continuously rotating half-wave plate on the POLARBEAR telescope*, *JCAP* **05** (2017) 008 [`arXiv:1702.07111`] [ɪɴSPIRE].

[83] R.J. Thornton et al., *The Atacama Cosmology Telescope: The polarization-sensitive ACTPol instrument*, *Astrophys. J. Suppl.* **227** (2016) 21 [`arXiv:1605.06569`] [ɪɴSPIRE].

[84] P. Virtanen et al., *SciPy 1.0–Fundamental Algorithms for Scientific Computing in Python*, *Nature Meth.* **17** (2020) 261 [`arXiv:1907.10121`] [ɪɴSPIRE].

[85] Y. Wiaux, J.D. McEwen, P. Vandergheynst and O. Blanc, *Exact reconstruction with directional wavelets on the sphere*, *Mon. Not. Roy. Astron. Soc.* **388** (2008) 770 [`arXiv:0712.3519`] [ɪɴSPIRE].

[86] M. Zaldarriaga and U. Seljak, *An all sky analysis of polarization in the microwave background*, *Phys. Rev. D* **55** (1997) 1830 [`astro-ph/9609170`] [ɪɴSPIRE].

[87] A. Zonca et al., *healpy: equal area pixelization and spherical harmonics transforms for data on the sphere in Python*, *J. Open Source Softw.* **4** (2019) 1298 [ɪɴSPIRE].

The Atacama Cosmology Telescop collaboration

Zachary Atkins [a,*], Adriaan J. Duivenvoorden [a,b], William R. Coulton [b], Frank J. Qu [c],
Simone Aiola [b], Erminia Calabrese [d], Grace E. Chesmore [e], Steve K. Choi [f,g],
Mark J. Devlin [h], Jo Dunkley [a,i], Carlos Hervías-Caimapo [j], Yilun Guan [k],
Adrien La Posta [l], Zack Li [m], Thibaut Louis [l], Mathew S. Madhavacheril [h],
Kavilan Moodley [n,o], Sigurd Naess [p], Federico Nati [q], Michael D. Niemack [f,g], Lyman Page [a],
Roberto Puddu [j], Maria Salatino [r,s], Cristóbal Sifón [t], Suzanne T. Staggs [a], Cristian Vargas [j],
Eve M. Vavagiakis [f], Edward J. Wollack [u]

[a] *Joseph Henry Laboratories of Physics, Jadwin Hall, Princeton University, Princeton, NJ 08544, U.S.A.*
[b] *Center for Computational Astrophysics, Flatiron Institute, New York, NY 10010, U.S.A.*
[c] *DAMTP, Centre for Mathematical Sciences, University of Cambridge,*
Wilberforce Road, Cambridge CB3 OWA, U.K.
[d] *School of Physics and Astronomy, Cardiff University, The Parade, Cardiff, Wales CF24 3AA, U.K.*
[e] *Department of Physics, University of Chicago, 5720 South Ellis Avenue, Chicago, IL 60637, U.S.A.*
[f] *Department of Physics, Cornell University, Ithaca, NY 14853, U.S.A.*
[g] *Department of Astronomy, Cornell University, Ithaca, NY 14853, U.S.A.*
[h] *Department of Physics and Astronomy, University of Pennsylvania,*
209 South 33rd Street, Philadelphia, PA 19104, U.S.A.
[i] *Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, U.S.A.*
[j] *Instituto de Astrofísica and Centro de Astro-Ingeniería, Facultad de Física, Pontificia Universidad*
Católica de Chile, Av. Vicuña Mackenna 4860, 7820436 Macul, Santiago, Chile
[k] *Dunlap Institute for Astronomy and Astrophysics, University of Toronto,*
50 St. George St., Toronto, ON M5S 3H4, Canada
[l] *Université Paris-Saclay, CNRS/IN2P3, IJCLab, 91405 Orsay, France*
[m] *Canadian Institute for Theoretical Astrophysics, University of Toronto, Toronto, ON M5S 3H8, Canada*
[n] *Astrophysics Research Centre, University of KwaZulu-Natal,*
Westville Campus, Durban 4041, South Africa
[o] *School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal,*
Westville Campus, Durban 4041, South Africa
[p] *Institute of Theoretical Astrophysics, University of Oslo, Oslo, Norway*
[q] *Department of Physics, University of Milano-Bicocca, Piazza della Scienza 3, 20126 Milano (MI), Italy*
[r] *Department of Physics, Stanford University, Stanford, California 94305, U.S.A.*
[s] *Kavli Institute for Astroparticle Physics and Cosmology, Stanford, California 94305, U.S.A.*
[t] *Instituto de Física, Pontificia Universidad Católica de Valparaíso, Casilla 4059, Valparaíso, Chile*
[u] *NASA Goddard Space Flight Center, 8800 Greenbelt Rd, Greenbelt, MD 20771, U.S.A.*

[*] *Corresponding author*