

# Detecting ultra-high-energy cosmic ray anisotropies through harmonic cross-correlations

Federico R. Urban<sup>1</sup>, Stefano Camera<sup>2,3</sup>, and David Alonso<sup>4</sup>

<sup>1</sup> CEICO, Institute of Physics of the Czech Academy of Sciences, Na Slovance 1999/2, 182 21 Prague, Czech Republic  
e-mail: federico.urban@fzu.cz

<sup>2</sup> Dipartimento di Fisica, Università degli Studi di Torino, Via P. Giuria 1, 10125 Torino, Italy

<sup>3</sup> INFN – Istituto Nazionale di Fisica Nucleare, Sezione di Torino, Via P. Giuria 1, 10125 Torino, Italy

<sup>4</sup> Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK

Received 21 May 2020 / Accepted 2 May 2021

## ABSTRACT

We propose an observable for ultra-high-energy cosmic ray (UHECR) physics: the harmonic-space cross-correlation power spectrum between the arrival directions of UHECRs and the large-scale cosmic structure mapped by galaxies. This cross-correlation has not yet been considered in the literature, and it permits a direct theoretical modelling of the main astrophysical components. We describe the expected form of the cross-correlation and show how, if the distribution of UHECR sources traces the large-scale cosmic structure, it could be easier to detect with current data than the UHECR auto-correlation. Moreover, the cross-correlation is more sensitive to UHECR anisotropies on smaller angular scales, it is more robust to systematic uncertainties, and it could be used to determine the redshift distribution of UHECR sources, making it a valuable tool for determining their origins and properties.

**Key words.** astroparticle physics – large-scale structure of Universe

## 1. Introduction

Ultra-high-energy cosmic rays (UHECRs), which impact the atmosphere of the Earth with energies in excess of 1 EeV ( $10^{18}$  eV), have remained a mystery since their discovery 59 yr ago (Linsley et al. 1961; Alves Batista et al. 2019). We do not know what they are: Observational data cannot yet distinguish between several variants of pure and mixed primary compositions (Castellina 2020; Bergman 2019). We do not know where they come from: The astrophysical sources that generate and accelerate UHECRs have not yet been identified, nor has the type of acceleration mechanism responsible for their formidable energies been discovered (Kotera & Olinto 2011).

What we do know is that the highest-energy rays are most likely extra-Galactic. First, if UHECRs were produced within the Galaxy, their arrival directions in the sky would be very different from what we observe (Tinyakov et al. 2016; Abbasi et al. 2017; Aab et al. 2017). Second, barring a cosmic conspiracy that puts an end to the injection spectrum at that very energy, UHECR interactions with cosmological background photons produce a sharp cutoff (the Greisen-Zatsepin-Kuzmin limit) in the spectrum corresponding to  $\sim 60$  EeV (Greisen 1966; Zatsepin & Kuzmin 1966), and a cutoff is indeed observed in the data (Abbasi et al. 2008; Abraham et al. 2008).

If the sources of UHECRs are extra-Galactic, they most probably correlate with the large-scale distribution of matter (large-scale structure, or LSS). The interactions with the background cold photons limit UHECR propagation to roughly a few hundred megaparsecs (for a review, see Kotera & Olinto 2011). Therefore, the UHECR flux distribution in the sky should be to some extent anisotropic since below 100 Mpc, roughly comparable with the scale of homogeneity expected in the standard

cosmological model, the LSS is anisotropic (Pan & Coles 2000; Scrimgeour et al. 2012; Alonso et al. 2015).

How the anisotropy of UHECR sources manifests itself in the observed flux on Earth therefore depends on the original anisotropy of the sources, the UHECR chemical composition, and the properties of intervening magnetic fields – Galactic (GMF) and extra-Galactic (xGMF) – that deflect UHECRs and distort the original anisotropic patterns. Chemical composition and magnetic fields are degenerate when it comes to UHECR deflections since such deflections depend on  $ZB/E$ , where  $Z$  is the atomic number,  $B$  is the strength of the magnetic field, and  $E$  is the UHECR energy: doubling the field strength is equivalent to doubling the charge (or halving the energy). Chemical composition instead is the only factor that determines the UHECR propagation length at a given energy: different nuclei come from different portions of the Universe and carry different anisotropic imprints, but the relationship between the two is non-monotonic and non-trivial (see e.g., d’Orfeuille et al. 2014; di Matteo & Tinyakov 2018).

To a large extent, the statistics of the anisotropies in the distribution of UHECRs can be characterised by the UHECR angular auto-correlation (AC), which, in harmonic space, takes the form of the angular power spectrum coefficients  $C_\ell$ . Here, the  $\ell$ -th multipole quantifies the variance of the anisotropies on angular scales  $\theta \sim \pi/\ell$  (Sommers 2001; Tinyakov & Urban 2015, see our Appendix A for further details). To date, the number of UHECRs collected at the highest energies is low – of the order of a hundred above the cutoff (Alves Batista et al. 2019). Because of this, the UHECR flux is dominated by Poisson statistics: the AC is mostly determined by shot noise, making the underlying correlation with the LSS very hard to detect. Indeed, the indications for anisotropy in the data are tenuous: save for

a low-energy dipole (Aab et al. 2017) and a high-energy hot spot (Abbasi et al. 2014), the angular distribution of UHECR arrival directions appears to be nearly isotropic (di Matteo et al. 2020). Moreover, no anisotropies have been detected at small scales,  $\ell \gtrsim 10$  (Deligny 2015; di Matteo et al. 2018), although there are hints at intermediate scales (Caccianiga 2020).

In this work we quantify the possibility of detecting the anisotropy in the UHECR flux through the harmonic-space power spectrum of the cross-correlation (XC) between UHECR counts and the distribution of galaxies. Such a XC technique was previously proposed to study the anisotropy of the  $\gamma$ -ray sky by Camera et al. (2013), Fornengo & Regis (2014), and Pinetti et al. (2019) and proved successful for several tracers of the LSS (Fornengo et al. 2015; Cuoco et al. 2015; Branchini et al. 2017; Ammazalorso et al. 2020). A search for a XC between UHECRs and high-energy photons was performed in Alvarez et al. (2016). If UHECR sources statistically trace the LSS, then the positions of these sources, and the arrival directions of UHECRs (if not strongly affected by intervening magnetic fields), should have a non-zero correlation with a galaxy sample up to a given distance. Therefore, the detection or non-detection of the XC signal with galaxies at different redshifts would allow us to test whether UHECR sources are distributed according to the LSS, and to quantify the extent to which the UHECR transfer function, determined by energy losses and intervening magnetic fields, is independent from direction.

There are at least three features that differentiate the XC from other methods (see for instance Koers & Tinyakov 2009 and references therein). First, systematic uncertainties of different ‘messengers’, or observables, should not cross-correlate, and, under some conditions, statistical noise should not strongly cross-correlate either. This is because different experiments are different machines exploiting different physical effects. However, within a single dataset, for instance the set of arrival directions of UHECRs, the AC of the noise and systematic errors for that set are certainly non-zero and contribute to hiding any underlying ‘true’ signal. Examples of these systematics for UHECRs would be perturbations in the arrival directions due to deflection by the GMF<sup>1</sup> or spatial fluctuations in the energy calibration that give rise to leakage from different energy bins<sup>2</sup>. Thus, in this sense the XC is an experimentally cleaner observable.

Second, in the limit where the UHECR sources are numerous but UHECR detections themselves are not, we can assume that we observe at most one UHECR per source (as seems to be the case given the lack of obvious UHECR multiplets; Abreu et al. 2012; Aab et al. 2019). The much higher number of galaxies leads to a significant improvement in the signal-to-noise ratio (S/N) of this XC (see the discussion in Sect. 3). This effectively allows us to probe the anisotropies on smaller scales through the XC than through the AC, underlying the importance of using both observables.

There are several reasons why those smaller scales ( $\ell > 10$ ) are interesting. The experimental angular resolution of UHECR events is around  $1^\circ$ , which corresponds to  $\ell \sim 200$ : From an experimental perspective, we are not fully taking advantage of the data we already have. Furthermore, small-scale power in the LSS angular distribution is comparable to that at large scales: if UHECRs bear the imprint of the LSS, this small-

scale power is not completely suppressed by the GMF, especially once the structured component of the GMF is taken into account (Dundović & Sigl 2019); moreover, the substructures of the GMF themselves imprint small-scale anisotropies in cosmic rays (CRs) of petaelectronvolt (PeV) energies (Giacinti & Sigl 2012), and it is possible that such structures can be present at higher energies. Lastly, small-scale anisotropies can be detected separately in different regions of the sky, allowing us to probe, for example, different GMF structures independently.

Third, while most analyses have looked at the real-space correlation between UHECRs and the LSS (Kashti & Waxman 2008; Oikonomou et al. 2013; Abreu et al. 2010; Aab et al. 2015; Takami et al. 2009), we express our results here in terms of harmonic-space power spectra. These are common observables in cosmological studies and are based on a natural decomposition of the celestial sphere. They also allow for a straightforward visualisation of the main components of the astrophysical model (radial kernels, details of the galaxy-matter connection), which is one of the main novel aspects of this work.

In this paper we introduce a formalism to model the AC and XC, and we apply it to a vanilla proton-only model for UHECR injection in order to quantify the differences between the two observables and the detectability of the anisotropies on different scales with existing experimental facilities. We defer to upcoming work: (i) a more detailed discussion of the dependence of the XC on UHECR injection and source properties, (ii) a realistic treatment of the UHECR experimental setup, such as non-uniform sky coverage, and (iii) a full treatment of the effects of the GMF and xGMF on the signal.

This paper is organised as follows. In Sect. 2 we introduce the formalism to describe the UHECR flux, the distribution of galaxies, and the AC and XC. We apply this formalism to a hypothetical UHECR dataset in Sect. 3, where we obtain and compare the AC and the XC. We summarise our findings and conclude with an outlook for future work in Sect. 4. Appendix A presents useful standard formulae pertaining to angular power spectra.

## 2. Theoretical model

### 2.1. UHECR flux

Let  $\mathcal{E}(E_{\text{inj}})$  be the (angle-integrated, isotropic) emissivity<sup>3</sup> of CRs for a given galaxy (number of CRs of energy  $E_{\text{inj}}$  emitted per unit energy, per unit time):

$$\mathcal{E}[E_{\text{inj}}] := \frac{dN_{\text{inj}}}{dE_{\text{inj}} dt_{\text{inj}}}. \quad (1)$$

The subscript ‘inj’ (injection) here indicates quantities evaluated in the rest frame of the emitting source. Due to the expansion of the Universe and to interactions between CRs and cosmic background light, the injected energy of a CR, whose energy at detection is  $E$ , is given by  $E_{\text{inj}}(E, z)$ , where  $z$  is the redshift of the source. In the absence of scattering processes, the energy losses are adiabatic:  $E_{\text{inj}} = (1 + z)E$ . The differential emissivity (i.e., per unit solid angle) is  $\epsilon := \mathcal{E}/4\pi$ , assuming isotropic emission. We parameterise the emissivity as a power law of energy:

$$\mathcal{E}[E_{\text{inj}}] \propto E_{\text{inj}}^{-\gamma}. \quad (2)$$

Energies here are always expressed in exaelectronvolts (EeV) for convenience.

<sup>3</sup> Our definition of emissivity differs from the one used in e.g., radio astronomy, which quantifies the energy (instead of the number) emitted per unit time, volume, and solid angle.

<sup>1</sup> Magnetic deflections can also affect the size of a dataset in the case of partial sky coverage (where UHECR events could migrate to and/or from the observed patch of the sky).

<sup>2</sup> Overall energy miscalibration would still need to be modelled and accounted for in the theory predictions.

The quantity measured on Earth is the observed number of events per unit time, energy interval, detector area, solid angle on the sky, and (assuming source redshifts can be measured) redshift interval. We can relate this number to the emissivity through

$$\frac{dN}{dE dt dA d\Omega dz} = \frac{n_{s,c} \mathcal{E}(E_{\text{inj}})}{4\pi (1+z) H(z)} \frac{dE_{\text{inj}}}{dE}, \quad (3)$$

where  $H(z)$  is the Hubble parameter and  $n_{s,c}$  is the volumetric number density of CR sources; we set  $c = 1$ , and we ignored sub-dominant light-cone and relativistic effects (Challinor & Lewis 2011; Bonvin & Durrer 2011).

We are interested in the number of UHECRs detected above a given energy threshold,  $E_{\text{cut}}$  (defined in the observer's frame), and integrated over source redshifts, from the direction  $\hat{n}$ :

$$\begin{aligned} \Phi(E_{\text{cut}}, \hat{n}) &:= \int_0^\infty dz \int_{E_{\text{cut}}}^\infty dE \frac{dN}{dE dt dA d\Omega dz} \\ &= \int \frac{dz}{(1+z)H(z)} \frac{n_{s,c}(z, \chi \hat{n})}{4\pi} \int_{E_{\text{cut}}}^\infty dE \frac{dE_{\text{inj}}}{dE} \mathcal{E}(E_{\text{inj}}), \end{aligned} \quad (4)$$

where  $\chi(z)$  is the radial co-moving distance.

We can write the number density of sources as  $n_{s,c}(z, \chi \hat{n}) = \bar{n}_{s,c}(z) [1 + \delta_s(z, \chi \hat{n})]$ , where  $\delta_s$  is the galaxy over-density. Assuming a non-evolving galaxy population, namely  $\bar{n}_{s,c}(z) = \bar{n}_{s,c}(0)$ , and a power-law UHECR spectrum (as in Eq. (2)), we obtain:

$$\Phi(E_{\text{cut}}, \hat{n}) \propto \frac{\bar{n}_{s,c}}{4\pi} \int \frac{d\chi}{(1+z)} \frac{E_{\text{inj}}^{1-\gamma}(E_{\text{cut}}, z)}{1-\gamma} [1 + \delta_s(z, \chi \hat{n})]. \quad (5)$$

### 2.1.1. Attenuation

The attenuation factor,  $\alpha(E_{\text{cut}}, z; \gamma, Z)$ , is defined as the number of events reaching the Earth with  $E > E_{\text{cut}}$ , divided by the number of events that would have reached the Earth if there were no energy losses at a given distance:

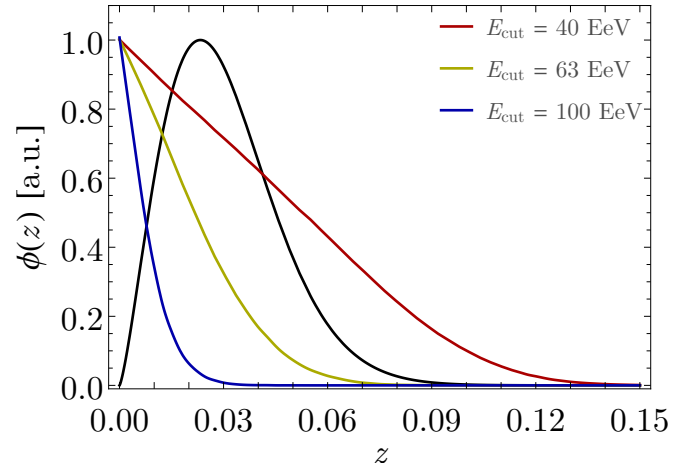
$$\alpha(z, E_{\text{cut}}; \gamma, Z) := \frac{E_{\text{inj}}^{1-\gamma}(E_{\text{cut}}, z)}{E_{\text{cut}}^{1-\gamma}}. \quad (6)$$

The attenuation,  $\alpha$ , is a function of the energy cut and redshift as well as of the injection spectral slope and chemical composition. In terms of  $\alpha$ , the direction-dependent integral flux is

$$\Phi(E_{\text{cut}}, \hat{n}) \propto \frac{\bar{n}_{s,c} E_{\text{cut}}^{1-\gamma}}{4\pi (1-\gamma)} \int d\chi \frac{\alpha(z, E_{\text{cut}})}{(1+z)} [1 + \delta_s(z, \chi \hat{n})]. \quad (7)$$

In this paper, to introduce the formalism, we chose to work with a toy proton-only model with injection slope  $\gamma = 2.6$  as in model (4) of d'Orfeuille et al. (2014) or model (i) of di Matteo & Tinyakov (2018). In order to obtain the attenuation factor for our injection model, we followed  $10^6$  events with *SimProp* v2r4 (Aloisio et al. 2017) with energies above  $E = 10$  EeV (with an upper cutoff of  $E = 10^5$  EeV), for redshifts up to  $z = 0.3$ , and counted the number of events reaching the Earth with  $E > E_{\text{cut}}$  for different values of  $E_{\text{cut}}$ . With *SimProp* we accounted for all energy losses, namely adiabatic losses and losses due to interactions with cosmic microwave background (CMB) photons and with extra-Galactic background photons according to the model in Stecker et al. (2006). The UHECR radial kernels, defined in the next section, obtained from the attenuation factor,  $\alpha$ , for different energies are shown in Fig. 1.

It should be noted that we assumed CR energy losses to be, to first order, isotropic, that is, we ignored angular anisotropies in



**Fig. 1.** Radial kernels for the two observables under consideration. The solid black line shows the approximate redshift distribution of galaxies in the 2MRS sample using the fit found by Ando et al. (2018). The red, yellow, and blue lines show the radial kernel for the UHECR flux (Eq. (10)) for the three energy thresholds studied here (40 EeV, 63 EeV, and 100 EeV, respectively).

the CMB and extra-Galactic background light, which are completely negligible for our analysis. Moreover, for simplicity here we work with full-sky uniform coverage, but the analysis can be readily generalised to non-uniform and partial sky coverage.

### 2.1.2. Anisotropies

We defined the anisotropy in the UHECR distribution as the over-density of rays detected as a function of sky position  $\hat{n}$  as

$$\Delta_{\text{CR}}(\hat{n}, E_{\text{cut}}) := \frac{\Phi(\hat{n}, E_{\text{cut}}) - \bar{\Phi}(E_{\text{cut}})}{\bar{\Phi}(E_{\text{cut}})}, \quad (8)$$

where  $\bar{\Phi}(E_{\text{cut}})$  is the sky-averaged UHECR flux. From the results in the previous section, this quantity is related to the three-dimensional over-density of UHECR sources  $\delta_s(z, \chi \hat{n})$  through

$$\Delta_{\text{CR}}(\hat{n}, E_{\text{cut}}) = \int d\chi \phi_{\text{CR}}(\chi) \delta_s(z, \chi \hat{n}), \quad (9)$$

where the UHECR radial kernel (or window function) is

$$\phi_{\text{CR}}(z) := \left[ \int d\tilde{z} \frac{\alpha(\tilde{z})}{H(\tilde{z})(1+\tilde{z})} \right]^{-1} \frac{\alpha(z)}{(1+z)}. \quad (10)$$

Figure 1 shows the radial kernels for UHECRs with energy thresholds  $E_{\text{cut}} = 40, 63,$  and  $100$  EeV; as expected, the lower the energy, the farther UHECRs propagate.

## 2.2. Galaxies

We considered the AC of the UHECR anisotropy, Eq. (9), and its XC with the galaxy number count fluctuations. In particular, we worked with the projected over-density of sources for a given galaxy sample,

$$\Delta_g(\hat{n}) := \frac{N_g(\hat{n}) - \bar{N}_g}{\bar{N}_g}, \quad (11)$$

where  $N_g(\hat{n})$  is the number of galaxies in a given direction ( $\hat{n}$ ) and  $\bar{N}_g$  is its average over the celestial sphere. This is related to

the three-dimensional galaxy over-density  $\delta_g(z, \chi \hat{\mathbf{n}})$  via

$$\Delta_g(\hat{\mathbf{n}}) = \int d\chi \phi_g(\chi) \delta_g(z, \chi \hat{\mathbf{n}}), \quad (12)$$

where  $\phi_g(\chi)$  is the weighted distribution of galaxy distances. In general, we have assumed that we have redshift information for all galaxies in the catalogue and that we can use that information to apply a distance-dependent weight,  $w(\chi)$ . In that case, the galaxy over-density kernel,  $\phi_g(\chi)$ , is given by

$$\phi_g(\chi) := \left[ \int d\tilde{\chi} \tilde{\chi}^2 w(\tilde{\chi}) \bar{n}_{g,c}(\tilde{\chi}) \right]^{-1} \chi^2 w(\chi) \bar{n}_{g,c}(\chi), \quad (13)$$

where  $\bar{n}_{g,c}$  is the co-moving number density of galaxies in the sample.

If no weights are applied – namely,  $w(\chi) = 1$  – then

$$\int d\chi \chi^2 w(\chi) \bar{n}_{g,c}(\chi) = \bar{N}_{\Omega,g}, \quad (14)$$

where  $\bar{N}_{\Omega,g}$  is the angular number density of galaxies (i.e., the number of galaxies per steradian).

Figure 1 shows the radial kernel for a low-redshift galaxy survey, modelled after the 2MASS Redshift Survey (2MRS; Huchra et al. 2012). This constitutes one of the most complete full-sky spectroscopic low-redshift surveys, and we use it as our fiducial galaxy sample in this paper. In this work we consider full-sky datasets for simplicity, but the generalisation of our results for an incomplete sky coverage is straightforward. In the case of a realistic setup based on 2MRS, a sky coverage around 70% will only degrade the signal by a factor of  $\sqrt{0.7} \simeq 0.86$ .

### 2.3. Power spectra

We are interested in detecting the intrinsic anisotropies in the distribution of UHECRs by considering the different two-point functions built from  $\Delta_{\text{CR}}$  and  $\Delta_g$ . A given observation of any of these fields will consist of both signal  $S$  and noise  $N$ :  $\Delta_a = s_a + n_a$  (where  $a, b \in \{\text{CR}, g\}$ ). Assuming signal and noise to be uncorrelated, the corresponding power spectra can be split into both components, namely

$$C_\ell := \mathcal{S}_\ell + \mathcal{N}_\ell, \quad (15)$$

where  $\mathcal{S}_\ell$  and  $\mathcal{N}_\ell$  are the power spectra of  $S$  and  $N$ , respectively. In our case, the signal is the intrinsic clustering of both UHECRs and galaxies due to the underlying LSS, while the noise is sourced by the discrete nature of both tracers as Poisson noise. A brief review of the mathematics behind angular power spectra is given in Appendix A.

#### 2.3.1. Signal power spectra

The angular power spectrum,  $\mathcal{S}_\ell^{ab}$ , between two projected quantities,  $\Delta_a$  and  $\Delta_b$ , is related to their three-dimensional power spectrum,  $P_{ab}(z, k)$ , by

$$\mathcal{S}_\ell^{ab} = \int \frac{d\chi}{\chi^2} \phi_a(\chi) \phi_b(\chi) P_{ab} \left( z, k = \frac{\ell + 1/2}{\chi} \right), \quad (16)$$

where  $\phi_a$  and  $\phi_b$  are the radial kernels of both quantities.

The final piece of information needed in order to estimate the expected AC and XC signals is the power spectrum of the three-dimensional over-densities  $\delta_s$  and  $\delta_g$ . In general, the clustering properties of galaxies and UHECR sources will depend

on the specifics of the relationship between galaxies and dark matter, as well as on the astrophysical properties of the UHECR sources. To simplify the discussion, here we have assumed that all UHECR sources are also galaxies of the 2MASS sample (i.e.,  $\delta_s = \delta_g$ ).

At this point, one might be tempted to use a linear bias prescription (Mo & White 1996) to relate the galaxy and matter power spectra. However, as we show in Sect. 3, since the UHECR radial kernel peaks at  $z = 0$  and covers only low redshifts, the CR flux AC probes mostly sub-halo scales for which a non-perturbative description of structure formation is necessary. To achieve this, we used here a halo model prescription (Peacock & Smith 2000) that is based on the halo occupation distribution model used by Ando et al. (2018) to describe the 2MRS sample. In this model, the galaxy power spectrum is given by two contributions,

$$P_{gg}(z, k) = P_{gg}^{1h}(z, k) + P_{gg}^{2h}(z, k), \quad (17)$$

which are the so-called one-halo and two-halo terms. The former dominates on small scales and describes the distribution of galaxies within the halo, while the latter is governed by the clustering properties of dark matter haloes. The halo occupation distribution is then based on a prescription to assign central and satellite galaxies to haloes of different masses. Although we summarise this model in Appendix C, we refer the reader to Ando et al. (2018) and references therein for further details about the specifics of the halo occupation distribution model used.

#### 2.3.2. Shot noise

Both projected over-densities,  $\Delta_{\text{CR}}$  and  $\Delta_g$ , are associated with discrete point processes, represented by the angular positions of the UHECRs and the galaxies in each sample. In this case, even in the absence of intrinsic correlations between the different fields, their power spectra receive a non-zero white contribution, given by

$$\mathcal{N}_\ell^{ab} = \frac{\bar{N}_{\Omega,a \wedge b}}{\bar{N}_{\Omega,a} \bar{N}_{\Omega,b}}, \quad (18)$$

where  $\bar{N}_{\Omega,a}$  ( $\bar{N}_{\Omega,b}$ ) is the angular number density of points in sample  $a$  or  $b$ , and  $\bar{N}_{\Omega,a \wedge b}$  is the angular number density of points shared in common. In our case this corresponds to the number of UHECRs originating from galaxies in the galaxy sample. For simplicity, we have assumed that the galaxy survey under consideration is sufficiently complete, such that all UHECRs are associated with an observed galaxy. In this case, the shot noise contributions to the power spectra are

$$\mathcal{N}_\ell^{\text{CR CR}} = (\bar{N}_{\Omega, \text{CR}})^{-1}, \quad (19)$$

$$\mathcal{N}_\ell^{gg} = \mathcal{N}_\ell^{g \text{ CR}} = (\bar{N}_{\Omega,g})^{-1}. \quad (20)$$

Since typically  $\bar{N}_{\Omega, \text{CR}} \ll \bar{N}_{\Omega,g}$ , then  $\mathcal{N}_\ell^{g \text{ CR}} \ll \mathcal{N}_\ell^{\text{CR CR}}$ , and therefore we neglect  $\mathcal{N}_\ell^{g \text{ CR}}$  in what follows. We have explicitly checked that the cross-noise can indeed be safely neglected in all our estimates and numerical results.

We note that when non-flat weights are applied to the galaxy catalogue, the resulting noise power spectrum reads

$$\mathcal{N}_\ell^{gg} = \frac{\int d\chi \chi^2 w^2(\chi) \bar{n}_{g,c}(\chi)}{\left[ \int d\chi \chi^2 w(\chi) \bar{n}_{g,c}(\chi) \right]^2}. \quad (21)$$

For  $w(\chi) = 1$ , Eq. (14) holds, and we recover the result in Eq. (19).

### 2.3.3. Optimal weights

We can use the results in this section to derive optimal weights  $w(\chi)$  to maximise the signal-to-noise of the galaxy-UHECR XC. We pixelised the celestial sphere and considered the UHECRs in a given pixel  $p$ ,  $\Phi_p$ , as well as the vector  $N_{p,i}$ , which contains the number of galaxies along the same pixel in intervals of distance  $\chi_i$ . The optimal weights  $w_i := w(\chi_i)$  can be found by maximising the likelihood of  $\Phi_p$  given  $N_{p,i}$  (Alonso et al. 2020) and are given by the so-called Wiener filter, namely

$$w_i = \sum_j \mathbf{Cov}^{-1}(N_{p,i}, N_{p,j}) \mathbf{Cov}(\Phi_p, N_{p,j}). \quad (22)$$

Here,  $\mathbf{Cov}(x, y)$  is the covariance matrix of two vectors,  $x$  and  $y$ .

Assuming Poisson statistics, we can use the results from the previous section to show (see Appendix B) that

$$w(\chi) = \frac{\alpha[z(\chi), E_{\text{cut}}, \gamma, Z]}{[1 + z(\chi)]\chi^2 \bar{n}_{g,c}(\chi)}. \quad (23)$$

In hindsight, this result is obvious: By inspecting Eqs. (10) and (13), we see that the optimal weights modify the radial galaxy kernel,  $\phi_g$ , to make it identical to the UHECR kernel,  $\phi_{\text{CR}}$ , thereby building the most likely estimate of the UHECR flux map from the galaxy positions. As we will see, this involves up-weighting galaxies at low redshifts, where the UHECRs that reach the Earth most likely originate, but few galaxies can be found due to volume effects. We notice that the weights are completely driven by the theoretical model for UHECR propagation and do not depend on the actual data (or their errors). We show how the use of optimal weights can improve the S/N for the XC in Sect. 3.2.

### 2.4. Intervening magnetic fields

The Milky Way is host to a magnetic field of a few  $\mu\text{G}$  (Boulanger et al. 2018), which is the screen that befores UHECR sources. The variety of the parametric models of the GMF – which disagree on the GMF functional forms and parameters, particularly on GMF substructures – reflects the complexity of the GMF, and, at the moment, none can be taken at face value (Boulanger et al. 2018; Unger & Farrar 2017). As a guideline, we can expect the GMF to deflect a UHECR with energy  $E = 100 \text{ EeV}$  by a few degrees for most parts of the sky, except for certain directions close to the Galactic plane (d’Orfeuille et al. 2014; see also Pshirkov et al. 2013). UHECRs will also be affected by any intervening xGMF, whose strength, shape, and filling factors vary by several orders of magnitude for different models and estimates (Subramanian 2016); however, the xGMF is believed to have a sub-dominant effect on large-scale UHECR propagation (Pshirkov et al. 2016).

Simple prescriptions to account for part of the effects related to the GMF and the xGMF include smearing the map of sources below a certain angular scale or mixing that map with an isotropic one (similar to what is done to take into account catalogue incompleteness beyond a certain distance; see e.g., Koers & Tinyakov 2009). Smearing the source map in our language is as simple as introducing a (Gaussian) beam in the signals as

$$\mathcal{S}_\ell^{gg} \rightarrow \mathcal{S}_\ell^{gg} \mathcal{B}_\ell^2, \quad \mathcal{S}_\ell^{\text{gCR}} \rightarrow \mathcal{S}_\ell^{\text{gCR}} \mathcal{B}_\ell \quad (24)$$

for the AC and XC, respectively, where

$$\mathcal{B}_\ell := \frac{1}{\pi \theta_{\text{smear}}^2} \exp[-\ell(\ell+1)\theta_{\text{smear}}^2] \quad (25)$$

and  $\theta_{\text{smear}}$  is the smearing angle.

However, these solutions tend to be rather artificial and inaccurately destroy potential signals or structures in the spectra that we are examining. More precisely, we know that the largest effects due to the small-scale GMF are not isotropic and in fact vary quite considerably across the sky (see e.g., Pshirkov et al. 2013; di Matteo & Tinyakov 2018). More precisely, Pshirkov et al. (2013) found that the UHECR deflections are majorated by the function

$$\theta_{\text{smear}}(b) \leq \left( \frac{40 \text{ EV}}{E/Z} \right) \frac{1^\circ}{\sin^2 b + 0.15}, \quad (26)$$

where  $b$  is elevation. Therefore, if we smear the whole sky with the same smearing angle, we are not faithfully representing the sky, and, depending on the smearing angle, we either underestimate the deflections in certain regions or overestimate them in other regions (or both). Moreover, these solutions do not account for the large-scale galactic field, which is the dominant effect and cannot be described by a simple smearing.

For these reasons, and in order to best introduce the method, in this first theoretical work we take a pragmatic approach and, keeping in mind all the caveats listed above, only briefly discuss the effect of a (constant) smearing angle on the AC and XC (see also Appendix D). We neglect all other effects of intervening magnetic fields.

## 3. Results

### 3.1. Signal-to-noise ratio

We estimate the S/N of the UHECR anisotropies as the square root of the Fisher matrix element corresponding to an effective amplitude parameter,  $A_{\text{CR}}$ , multiplying the signal component of  $\Delta_{\text{CR}}$  with a fiducial value  $A_{\text{CR}} = 1$  (Heavens 2009), namely

$$S/N^2 := \sum_{\ell=\ell_{\text{min}}}^{\ell_{\text{max}}} \left( \frac{\partial \mathcal{S}_\ell}{\partial A_{\text{CR}}} \right)^\top \mathbf{Cov}_{\ell\ell'}^{-1} \frac{\partial \mathcal{S}_{\ell'}}{\partial A_{\text{CR}}}, \quad (27)$$

$$= \sum_{\ell=\ell_{\text{min}}}^{\ell_{\text{max}}} (S/N_\ell)^2, \quad (28)$$

where  $\mathcal{S}_\ell$  is a vector containing the signal contribution to the power spectra under consideration,  $\mathbf{Cov}$  is the covariance matrix of those power spectra, and  $S/N_\ell$  is the S/N of a single  $\ell$  mode. If the fields being correlated are Gaussian ( $\Delta_{\text{CR}}$ ,  $\Delta_g$  in our case), the covariance matrix can be estimated using Wick’s theorem to be

$$\mathbf{Cov}(C_\ell^{ab}, C_\ell^{cd}) = \frac{C_\ell^{ac} C_\ell^{bd} + C_\ell^{ad} C_\ell^{bc}}{(2\ell+1)\Delta\ell} \delta_{\ell\ell'}, \quad (29)$$

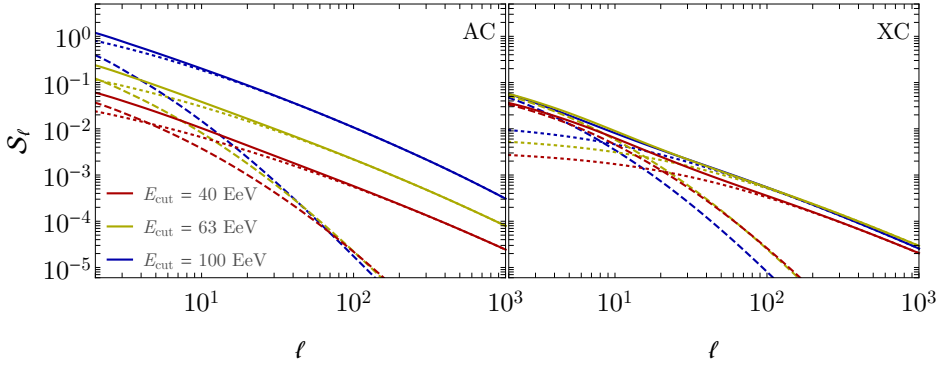
where  $\Delta\ell$  is the size of the multipole bin.

At this point we consider three different cases. The first is AC only. In this case we only have a measurement of the UHECR AC,  $C_\ell^{\text{CR CR}}$ . The S/N is given by

$$S/N^{\text{CR CR}} = \sqrt{\sum_{\ell=\ell_{\text{min}}}^{\ell_{\text{max}}} 2(2\ell+1) \left( \frac{\mathcal{S}_\ell^{\text{CR CR}}}{\mathcal{S}_\ell^{\text{CR CR}} + \mathcal{N}_\ell^{\text{CR CR}}} \right)^2}. \quad (30)$$

The second is XC only. In this case we only use the XC,  $C_\ell^{\text{gCR}}$ . The S/N is given by

$$S/N^{\text{gCR}} = \sqrt{\sum_{\ell=\ell_{\text{min}}}^{\ell_{\text{max}}} (2\ell+1) \frac{(\mathcal{S}_\ell^{\text{gCR}})^2}{C_\ell^{\text{gCR}} C_\ell^{\text{CR CR}} + (C_\ell^{\text{gCR}})^2}}. \quad (31)$$



**Fig. 2.** Angular AC and XC power spectra considered in this work. Dotted and dashed curves respectively refer to the one- and two-halo contribution to the total signal (solid curves).

In the third case we use all available data, that is, a data vector  $\mathcal{S}_\ell = (\mathcal{S}_\ell^{\text{CR CR}}, \mathcal{S}_\ell^{\text{g CR}}, \mathcal{S}_\ell^{\text{g g}})$ . Although this is the manifestly optimal scenario, XCs are arguably safer than ACs in terms of systematic errors, and therefore it is interesting to quantify the loss of information if only XCs are used.

Studying these three cases allows us to explore the benefits of using XCs versus ACs, as well as the relative amount of information in each of the different two-point functions. Given the relatively small number of UHECRs currently measured, shot noise in the UHECR flux is the dominant contribution to the uncertainties. Comparing Eqs. (30) and (31), we can see that the S/N scales as  $N_{\text{CR}}^{-1}$  and  $N_{\text{CR}}^{-1/2}$  for cases 1 and 2, respectively, highlighting the potential of XCs for achieving a detection.

### 3.2. Power spectra and signal-to-noise

The energy at which we chose to cut the UHECR integral spectrum,  $E_{\text{cut}}$ , determines the UHECR propagation horizon, which in turns determines the strength of the anisotropy. Moreover, the choice of  $E_{\text{cut}}$  for a given UHECR spectrum also determines the number of UHECR events we have to sample the anisotropic angular distribution. We expect a tradeoff between the two. At low energies the UHECR sample contains many more events than at high energies because the UHECR spectrum drops very steeply with energy. However, for the range of energies we are interested in, the galaxy sample is much larger; as such, this does not have as strong an effect for the XC as it does for the traditional AC (whose noise is determined by the number of UHECR events). Moreover, at low energies UHECRs propagate farther, and the larger line-of-sight averaging can dilute the expected anisotropy. Lastly, the effects of intervening magnetic fields are stronger – this is expected to have a significant impact on the anisotropies, albeit less so for the XC compared to the AC thanks to its stability against systematics. At high energies the UHECR horizon is smaller, UHECRs undergo smaller deflections, and the anisotropy should be more pronounced, but the number of events drops dramatically.

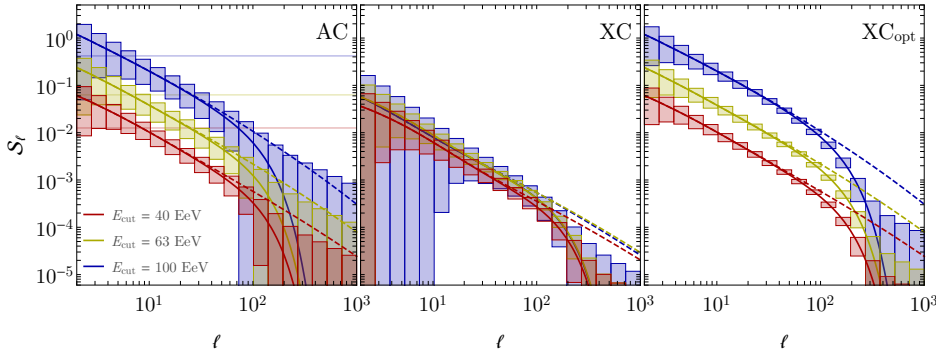
In order to determine at which energy we have the best chances of detecting the XC, we chose to work with three energy cuts at:  $E_{\text{cut}} = 10^{19.6} \text{ eV} \simeq 40 \text{ EeV}$ ,  $E_{\text{cut}} = 10^{19.8} \text{ eV} \simeq 63 \text{ EeV}$ , and  $E_{\text{cut}} = 10^{20} \text{ eV} = 100 \text{ EeV}$ . In a realistic scenario, based on currently available data (Alves Batista et al. 2019), we can expect to have about  $N_{\text{CR}} = 1000$ ,  $N_{\text{CR}} = 200$ , and  $N_{\text{CR}} = 30$  over the full sky for these three energy cuts, respectively. While this does not fully reflect a realistic situation – mostly because of magnetic deflections, which we do not take into account, and because current experimental facilities are limited in their field of view – our results nonetheless present a fair comparison between the two measures (AC and XC). This is owing to the

fact that all the salient information regarding UHECR datasets is represented in our estimates, namely the energy cut and with it all the UHECR energy losses, the available or expected number of events at that energy, and the angular resolution representative of what current experiments can do.

In Fig. 2 we show the expected signal for the AC (left panel) and the XC (right panel). Colours refer to the three energy cuts discussed above, namely red for  $E_{\text{cut}} \simeq 40 \text{ EeV}$ , yellow for  $E_{\text{cut}} \simeq 63 \text{ EeV}$ , and blue for  $E_{\text{cut}} = 100 \text{ EeV}$ . The dashed and dotted curves show the one-halo and two-halo contributions to the total power spectrum, with the sum of both shown by the solid curves. For simplicity, we have not included any beam smoothing in the plot. We can see how the signal for the XC is lower than for the AC, as is expected from the fact that the XC mixes two different radial kernels. If we employed optimal weights for the XC, the signal would become identical to that of the AC. In our simplistic linear treatment of perturbations, this happens because the UHECR and galaxy kernels would be identical. The statistical uncertainties for both correlation functions, however, would be different given their different shot noise levels.

To better understand the role of the different uncertainties on the theoretical signal, in Fig. 3 we again show the expected signal as in Fig. 2 (solid curves, same colour code) and include a  $1^\circ$  Gaussian smoothing beam to account for the angular resolution of UHECR experiments (for reference, we also show the beam-free prediction as dashed lines). On top of it, we present the corresponding  $\ell$ -binned  $1\sigma$  error bars as shaded boxes for 20 log-space multipole bins between  $\ell_{\text{min}} = 2$  and  $\ell_{\text{max}} = 1000$ . If we compare the leftmost and central panels, namely AC versus XC, it is easy to see how the range of multipoles where error bars are small enough to allow a detection is larger for XCs than for ACs for the  $E_{\text{cut}} \simeq 40 \text{ EeV}$  and  $E_{\text{cut}} \simeq 63 \text{ EeV}$  cases. However, for the sparser UHECR sample with  $E_{\text{cut}} = 100 \text{ EeV}$ , the opposite applies; more precisely, the detectable range of multipoles for the XC is smaller and pushed towards higher  $\ell$  compared to the AC. This is due to a combination of two factors: for the higher end of UHECR energies, the propagation horizon of UHECRs is small and the UHECR sky looks more anisotropic, boosting the AC. At the same time, the mismatch in kernels is prominent, the more so the higher the energy, and this drives the XC signal down. Combined with the larger shot noise in the UHECR data, this can explain the performance of the 100 EeV case – indeed, the UHECR shot noise is the main factor that prevents a detection of the signal at mid- $\ell$  values (the per- $\ell$  signal is  $1\sigma$  compatible with zero; see below).

In the rightmost panel of Fig. 3 we show the XC signal when we apply theoretical optimal weights. In this case the highest-energy set performs the best, and this is expected from the



**Fig. 3.** Expected power spectra and  $\ell$ -binned  $1\sigma$  uncertainties (shaded boxes), including a  $1^\circ$  Gaussian smoothing beam to account for the angular resolution of UHECR experiments (solid curves). For reference, horizontal lines in the leftmost plots denote shot noise levels, and the dashed curves show the beam-free prediction.

previous arguments: the signal is boosted back up to the same level of the AC because the kernels of galaxies and UHECRs now coincide. Additionally, while the uncertainty increases with energy as both samples become sparser, it is not large enough to hide the XC signal. It is worth noticing that the increase in galaxy power that we expect towards lower redshifts is significantly less relevant than the matching of the radial kernels.

In practice, using optimal weights may not be possible given the uncertainties in the radial kernel for UHECRs (we do not yet know the actual injection spectrum). The availability of redshift information in the galaxy catalogue, however, would allow us to turn this into an advantage: the UHECR kernel could be reconstructed by modifying the galaxy weights to maximise the signal-to-noise, essentially following the ‘clustering redshifts’ method used to reconstruct unknown redshift distributions in weak lensing data (Newman 2008).

To quantify the improvement in detectability brought by the XC, in Fig. 4 we present the cumulative  $S/N$  for all the data combinations discussed in Sect. 3.1, viz. AC alone (leftmost panel), XC alone (central panel), and all the data combined into a single data vector  $S_\ell$  (rightmost panel). In each panel, the left half shows the cumulative  $S/N$  as a function of the maximum multipole,  $\ell_{\max}$ , whilst the right half is for the cumulative  $S/N$  as a function of the minimum multipole,  $\ell_{\min}$ . In both cases, the case with all the data combined has, unsurprisingly, the largest  $S/N$ , but the contributions from AC and XC come from different angular scales. These scales in turn are sensitive to different redshift ranges depending on  $E_{\text{cut}}$ , which sets the propagation depth for UHECRs. This highlights the complementarity of the two observables.

The aforementioned sensitivity to different angular scales can be captured better by looking at Fig. 5, where we show the contribution to the total  $S/N$  from each integer multipole,  $S/N_\ell$ . The colour code is the same as throughout the paper, and we mark with horizontal dashed lines the thresholds corresponding to 1, 2, and  $3\sigma$  evidence for a one-parameter amplitude fit. These panels can be interpreted as evidence for anisotropy on a given scale; it is clear that the XC with galaxies helps to push the detectability of the signal to smaller scales (i.e. larger  $\ell$  values). This per- $\ell$   $S/N_\ell$  is a useful quantity for assessing whether the AC or the XC is the best observable for detecting the anisotropy in UHECRs, assuming that UHECRs trace the LSS.

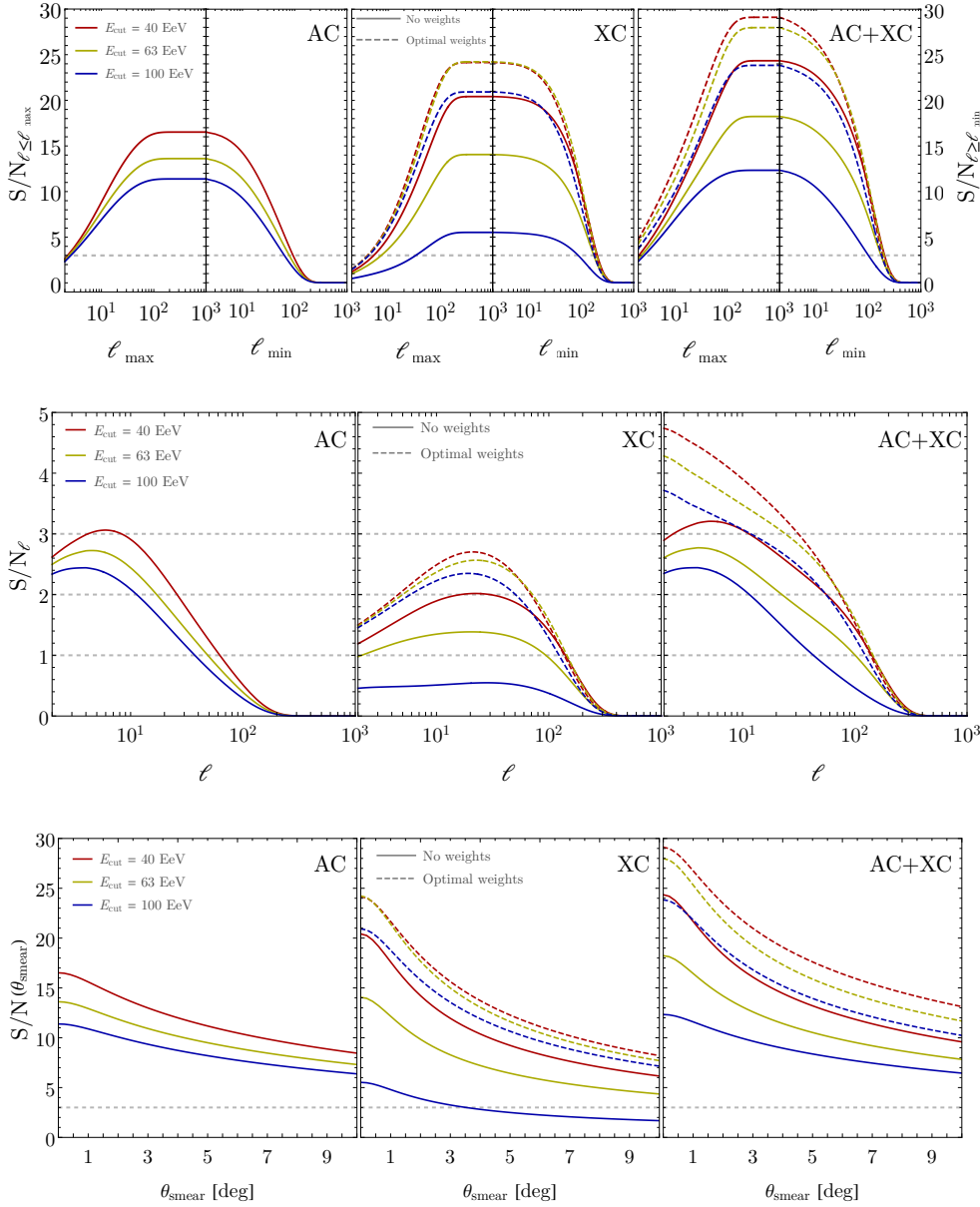
The sensitivity of the XC to small-scale anisotropies can be precious in realistic situations for two further reasons. First, a single Earth-based experiment is blind to a large fraction of the sky (roughly speaking, one celestial hemisphere); galaxy catalogues can also have incomplete sky coverage. Moreover, it might be advantageous (see our discussion of the direction-dependent magnetic deflections in Sect. 2.4) to restrict the UHECR dataset to a portion of the sky to maximise the chances

for a clean detection. In all these situations, the low harmonic multipoles are the most affected by these sky cuts. Second, if two experiments join their datasets, as the Telescope Array and Pierre Auger collaborations have done in their harmonic AC analysis, they need to cross-calibrate their sets, and this cross-calibration introduces errors that are significantly larger for low multipoles than for high multipoles (Abu-Zayyad et al. 2013; Aab et al. 2014; Deligny 2015; di Matteo et al. 2020).

As we argued in Sect. 2.4, there is no shortcut to account for the effects of the GMF on the AC and XC. Nonetheless, it is instructive to look at how the signal degrades with a simple smearing of the source map. To this end, we have plotted in Fig. 6 the total  $S/N$  for  $\ell = [2, 1000]$  as a function of the smearing angle,  $\theta_{\text{smear}}$ , of the galaxy map for the same energy cuts we have used so far. According to Eq. (26), the deflections for 40 EeV rigidity peak at around  $7^\circ$  near the Galactic centre, whereas more than half of the sky would be well described by a  $2.5^\circ$  smearing. It should be noted that, as mentioned in the introduction, the composition of UHECRs at the highest energies is unknown (Castellina 2020; Bergman 2019), and a heavier composition would imply lower rigidities and larger deflections. The smearing impacts the high-multipole regions in the XC more than it does in the AC, as expected, and degrades the XC more prominently at larger smearing angles (see also Appendix D).

The XC is strongly dependent on energy and the choice of weights. This means that we can disentangle small-scale anisotropies caused by the propagation through the GMF from the intrinsic anisotropies inherited from the LSS. In particular, any GMF-induced signal is erased at higher energies because UHECRs go straighter, whereas the LSS-inherited signal is enhanced because of the smaller propagation horizon. Moreover, any GMF-induced signal is indifferent to the weights we apply, whereas the signal from LSS anisotropies is strongly enhanced with the use of optimal weights.

Before closing this section, we remark that in a real experiment there will be modelled and un-modelled systematic errors to take into account. Systematic errors are expected to contribute to the AC more significantly than to the XC, particularly on large scales (low- $\ell$  end), for example the cross-calibration of two UHECR datasets. On the other hand, biased redshift information in galaxy catalogues or UHECR injection properties will affect both observables. To be clear, while the galaxy catalogue and the optimal weights, which depend on UHECR data for the reconstruction of the injection properties, do not enter into the AC obtained from UHECR data alone, they are needed once we test the source model (e.g., so that UHECRs correlate with the LSS). Hence, once systematic effects are taken into account, the  $S/N$  for the AC may decrease more than the  $S/N$  for the XC. This is one further motivation to explore the possibilities of and improvements from the use of XCs in UHECR anisotropy studies.



**Fig. 4.** S/N for UHECR flux anisotropies from different combinations of data, namely UHECR AC in the *leftmost panel*, XC in the *central panel*, and the combination of all data in the *rightmost panel*. In each panel, the *left half* shows the cumulative S/N as a function of the maximum multipole,  $\ell_{\max}$ , whereas the *right half* is for the cumulative S/N as a function of the minimum multipole,  $\ell_{\min}$ . The horizontal dashed line marks the  $3\sigma$  threshold for detection.

**Fig. 5.** S/N per multipole,  $S/N_{\ell}$ , for the AC signal, the XC signal with both normal and optimal weights, and their combination, AC+XC (*leftmost, central, and rightmost panel*, respectively). Different colours refer to different energy cuts, and the three horizontal dashed lines show the thresholds for 1, 2, and  $3\sigma$  detection.

**Fig. 6.** Total S/N,  $\sqrt{\sum_{\ell=\ell_{\min}}^{\ell_{\max}} S/N_{\ell}^2}$ , as a function of the smearing angle for the AC signal, the XC signal with both normal and optimal weights, and their combination, AC+XC (*leftmost, central, and rightmost panel*, respectively). Different colours refer to different energy cuts, and the horizontal dashed line shows the thresholds for  $3\sigma$  detection.

#### 4. Conclusions and outlook

In this work we have introduced a new observable for UHECR physics: the harmonic-space XC between the arrival directions of UHECRs and the distribution of the cosmic LSS as mapped by galaxies (Eq. (16)). We have developed the main theoretical tools necessary to model the signal and its uncertainties.

The take-away points of this study are as follows.

- The XC can be easier to detect than the UHECR AC for a range of energies and multipoles (see Figs. 3 and 5). This performance is mostly driven by the sheer number of galaxies that can trace the underlying LSS distribution, which is assumed to be the baseline distribution for both the UHECR flux and the galaxy angular distribution.
- The XC is more sensitive to small-scale angular anisotropies than the AC; conversely, the AC is more sensitive to large-scale anisotropies. This finding could therefore be instrumental in understanding properties of UHECR sources that would not be accessible otherwise.
- It is in principle possible to optimise the XC signal by assigning optimal redshift-dependent weights to sources in the

galaxy catalogue in order to match the UHECR radial kernel as determined by UHECR energy losses. Since matching the kernels has a strong impact on the XC, it could be possible to use this effect to reverse engineer the injection model (which defines the radial kernel).

- The great disruptor of UHECR anisotropies is the GMF. The XC, with its higher S/N and sensitivity to small angular scales, could be very useful in understanding the properties of the GMF (although we have not explored this angle here). Moreover, it may be possible, in the near future, to exploit a tomographic approach to disentangle the effects of intervening magnetic fields from different injection spectra and study different regions of the sky separately.

In our treatment we do not take any experimental uncertainties into account, with the exception of the experimental UHECR angular resolution. Moreover, we limit ourselves to a proton-only injection model and do not include the effects of the intervening magnetic fields. This choice was made in order to underline the physics behind our proposal and method. This method can be readily generalised and extended to include the

(theoretical and experimental) properties of the different galaxy and UHECR catalogues and the different injection models as well as to separate the number of events and energy cuts in order to best forecast the possibilities of present and upcoming UHECR datasets.

Moreover, in this initial work we have made the case for the XC between UHECRs and galaxies, but the logic and methods we have developed can be applied to other XCs with different matter tracers and different messengers. The distribution of visible matter in the sky can be traced not only by galaxies, but also by the thermal Sunyaev-Zeldovich effect. This effect is produced by the inverse Compton scattering of CMB photons by hot electrons along the line of sight. Because a thermal Sunyaev-Zeldovich map is a map of CMB photons, it is very accurate down to angles much smaller than a degree, and its signal peaks at low redshifts (Erler et al. 2018). This XC could therefore be useful in further disentangling the astrophysical properties of UHECR sources.

Charged UHECRs are not the only high-energy messengers whose production mechanisms and sources are unknown. The IceCube collaboration has recently detected a few high-energy astrophysical neutrinos with energies above 1 PeV (Aartsen et al. 2014). Such neutrinos are expected to be produced in the same extreme astrophysical sources as UHECRs and/or in their immediate surroundings. The XCs between neutrinos and the LSS will then inform us about the properties of the highest-energy astrophysical engines (see Fang et al. 2020 as well as Ando et al. 2015). Without the use of XCs (due to the very small number of neutrino events in current data), the detection of the anisotropic pattern could remain challenging into the foreseeable future (Fang et al. 2020; Sapienza 2020; Allison et al. 2016; Nelles 2019). Since neutrinos interact extremely weakly, they can propagate unhampered for long distances: their horizon is almost the entire visible Universe. Therefore, complementary information could be extracted from cross-correlating neutrinos with other tracers (in addition to galaxies), including CMB lensing (Planck Collaboration VIII 2020) and cosmic shear surveys (Mandelbaum 2018), both of which trace the overall matter distribution in the Universe, including both its dark and luminous components, out to higher redshifts with broader kernels (see Fornengo et al. 2015; Cuoco et al. 2015; Branchini et al. 2017; Ammazalorso et al. 2020 for the analogous analysis with  $\gamma$  rays). Measuring these XCs could reveal whether the most energetic particle accelerators in the Universe preferentially reside in high-density visible or dark environments.

*Acknowledgements.* FU wishes to thank A. di Matteo for useful correspondence, and P. Tinyakov for valuable comments on the manuscript. FU is supported by the European Regional Development Fund (ESIF/ERDF) and the Czech Ministry of Education, Youth and Sports (MEYS) through Project CoGraDS – CZ.02.1.01/0.0/0.0/15\_003/0000437. SC is supported by the Italian Ministry of Education, University and Research (MIUR) through Rita Levi Montalcini project ‘PROMETHEUS – Probing and Relating Observables with Multi-wavelength Experiments To Help Enlightening the Universe’s Structure’, and by the ‘Departments of Excellence 2018-2022’ Grant awarded by MIUR (L. 232/2016). DA acknowledges support from the Beecroft Trust, and from the Science and Technology Facilities Council through an Ernest Rutherford Fellowship, grant reference ST/P004474/1.

## References

Aab, A., Abreu, P., Aglietta, M., et al. 2014, *ApJ*, 794, 172  
Aab, A., Abreu, P., Aglietta, M., et al. 2015, *ApJ*, 804, 15

- Aab, A., Abreu, P., Aglietta, M., et al. 2017, *Science*, 357, 1266  
Aab, A., Abreu, P., Aglietta, M., et al. 2019, *ICRC*, C4, 20  
Aartsen, M. G., Ackermann, M., Adams, J., et al. 2014, *Phys. Rev. Lett.*, 113  
Abbasi, R. U., Abu-Zayyad, T., Allen, M., et al. 2008, *Phys. Rev. Lett.*, 100  
Abbasi, R. U., Abu-Zayyad, T., Allen, M., et al. 2014, *ApJ*, 790, L21  
Abbasi, R. U., Abu-Zayyad, T., Allen, M., et al. 2017, *Astropart. Phys.*, 86, 21  
Abraham, J., Abreu, P., Aglietta, M., et al. 2008, *Phys. Rev. Lett.*, 101  
Abreu, P., Aglietta, M., Ahn, E. J., et al. 2010, *Astropart. Phys.*, 34, 314  
Abreu, P., Aglietta, M., Ahn, E. J., et al. 2012, *Astropart. Phys.*, 35, 354  
Abu-Zayyad, T., Allen, M., Anderson, R., et al. 2013, *33rd International Cosmic Ray Conference*  
Allison, P., Bard, R., Beatty, J. J., et al. 2016, *Phys. Rev. D*, 93  
Aloisio, R., Boncioli, D., Di Matteo, A., et al. 2017, *JCAP*, 1711, 009  
Alonso, D., Salvador, A. I., Sánchez, F. J., et al. 2015, *MNRAS*, 449, 670  
Alonso, D., Cusin, G., Ferreira, P. G., & Pitrou, C. 2020, *Phys. Rev. D*, 102, 023002  
Alvarez, E., Cuoco, A., Mirabal, N., & Zaharijas, G. 2016, *JCAP*, 12, 023  
Alves Batista, R., Biteau, J., Bustamante, M., et al. 2019, *Front. Astron. Space Sci.*, 6, 23  
Ammazzalorso, S., Gruen, D., Regis, M., et al. 2020, *Phys. Rev. Lett.*, 124  
Ando, S., Tamborra, I., & Zandanel, F. 2015, *Phys. Rev. Lett.*, 115  
Ando, S., Benoit-Lévy, A., & Komatsu, E. 2018, *MNRAS*, 473, 4318  
Bergman, D. 2019, *PoS, ICRC2019*, 190  
Bonvin, C., & Durrer, R. 2011, *Phys. Rev. D*, 84  
Boulanger, F., Enßlin, T., Fletcher, A., et al. 2018, *JCAP*, 1808, 049  
Branchini, E., Camera, S., Cuoco, A., et al. 2017, *ApJS*, 228, 8  
Caccianiga, L. 2020, *PoS, ICRC2019*, 206  
Camera, S., Fornasa, M., Fornengo, N., & Regis, M. 2013, *ApJ*, 771, L5  
Castellina, A. 2020, *PoS, ICRC2019*, 004  
Challinor, A., & Lewis, A. 2011, *Phys. Rev. D*, 84  
Cooray, A., & Sheth, R. K. 2002, *Phys. Rept.*, 372, 1  
Cuoco, A., Xia, J.-Q., Regis, M., et al. 2015, *ApJS*, 221, 29  
Deligny, O. 2015, *PoS, ICRC2015*, 395  
di Matteo, A., & Tinyakov, P. 2018, *MNRAS*, 476, 715  
di Matteo, A., Deligny, O., Kawata, K., et al. 2018, *JPS Conf. Proc.*, 19  
di Matteo, A., Aloisio, R., Boncioli, D., et al. 2020, *PoS, ICRC2019*, 439  
d’Orfeuil, B. R., Allard, D., Lachaud, C., et al. 2014, *A&A*, 567, A81  
Dundović, A., & Sigl, G. 2019, *JCAP*, 1901, 018  
Erler, J., Basu, K., Chluba, J., & Bertoldi, F. 2018, *MNRAS*, 476, 3360  
Fang, K., Banerjee, A., Charles, E., & Omori, Y. 2020, *ApJ*, 894, 112  
Fornengo, N., & Regis, M. 2014, *Front. Phys.*, 2, 6  
Fornengo, N., Perotto, L., Regis, M., & Camera, S. 2015, *ApJ*, 802, L1  
Giacinti, G., & Sigl, G. 2012, *Phys. Rev. Lett.*, 109  
Greisen, K. 1966, *Phys. Rev. Lett.*, 16, 748  
Heavens, A. 2009, *Statistical Techniques in Cosmology*  
Huchra, J. P., Macri, L. M., Masters, K. L., et al. 2012, *ApJS*, 199, 26  
Kashfi, T., & Waxman, E. 2008, *JCAP*, 05, 006  
Koers, H. B. J., & Tinyakov, P. 2009, *JCAP*, 0904, 003  
Kotera, K., & Olinto, A. V. 2011, *ARA&A*, 49, 119  
Linsley, J., Scarsi, L., & Rossi, B. 1961, *Phys. Rev. Lett.*, 6, 485  
Mandelbaum, R. 2018, *ARA&A*, 56, 393  
Mo, H. J., & White, S. D. M. 1996, *MNRAS*, 282, 347  
Nelles, A. 2019, *EPJ Web Conf.*, 216, 01008  
Newman, J. A. 2008, *ApJ*, 684, 88  
Oikonomou, F., Connolly, A., Abdalla, F. B., et al. 2013, *JCAP*, 05, 015  
Pan, J., & Coles, P. 2000, *MNRAS*, 318, L51  
Peacock, J. A., & Smith, R. E. 2000, *MNRAS*, 318, 1144  
Pinetti, E., Camera, S., Fornengo, N., & Regis, M. 2020, *JCAP*, 07, 044  
Pshirkov, M. S., Tinyakov, P. G., & Urban, F. R. 2013, *MNRAS*, 436, 2326  
Planck Collaboration VIII. 2020, *A&A*, 641, A8  
Pshirkov, M. S., Tinyakov, P. G., & Urban, F. R. 2016, *Phys. Rev. Lett.*, 116  
Sapienza, P. 2020, *J. Phys. Conf. Ser.*, 1342  
Scrimgeour, M., Davis, T., Blake, C., et al. 2012, *MNRAS*, 425, 116  
Sommers, P. 2001, *Astropart. Phys.*, 14, 271  
Stecker, F. W., Malkan, M. A., & Scully, S. T. 2006, *ApJ*, 648, 774  
Subramanian, K. 2016, *Rept. Prog. Phys.*, 79  
Takami, H., Nishimichi, T., Yahata, K., & Sato, K. 2009, *JCAP*, 06, 031  
Tinyakov, P. G., & Urban, F. R. 2015, *J. Exp. Theor. Phys.*, 120, 533  
Tinyakov, P. G., Urban, F. R., Ivanov, D., Thomson, G. B., & Tirone, A. H. 2016, *MNRAS*, 460, 3479  
Unger, M., & Farrar, G. R. 2017, *35th Int. Cosmic Ray Conf.*, 35, 558  
Zatsepin, G. T., & Kuzmin, V. A. 1966, *JETP Lett.*, 4, 78 [Pisma Zh. Eksp. Teor. Fiz. 4,114(1966)]

## Appendix A: Power spectra

Three-dimensional fields  $\delta_a(\mathbf{x})$  can be decomposed into their Fourier modes

$$\delta_a(\mathbf{k}) := \int dk^3 \delta_a(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}}, \quad (\text{A.1})$$

whose covariance is the power spectrum  $P_{ab}(k)$ . Assuming statistical homogeneity and isotropy,  $P_{ab}(k)$  is implicitly defined by

$$\langle \delta_a(\mathbf{k}) \delta_b^*(\mathbf{k}') \rangle := \delta(\mathbf{k} - \mathbf{k}') P_{ab}(k), \quad (\text{A.2})$$

where the angle brackets denote averaging over ensemble realizations of the random fields inside them.

Equivalently, two-dimensional fields  $\Delta_a(\hat{\mathbf{n}})$  can be decomposed into their harmonic coefficients:

$$\Delta_{\ell m}^a := \int d\Omega Y_{\ell m}^*(\hat{\mathbf{n}}) \Delta_a(\hat{\mathbf{n}}), \quad (\text{A.3})$$

where  $\Omega = (\theta, \varphi)$  is the solid angle on the sky,  $Y_{\ell m}$  are the spherical harmonic functions,  $\hat{\mathbf{n}}$  is the line-of-sight direction, and in this work  $a = \{\text{CR}, g\}$ . The covariance of the  $\Delta_{\ell m}$  is the angular power spectrum,  $\mathcal{S}_{\ell}^{ab}$ , defined as

$$\langle \Delta_{\ell m}^a \Delta_{\ell' m'}^{b*} \rangle := \delta_{\ell\ell'} \delta_{mm'} \mathcal{S}_{\ell}^{ab}. \quad (\text{A.4})$$

For two projected fields,  $\Delta_a$  and  $\Delta_b$ , associated with three-dimensional fields  $\delta_a$  and  $\delta_b$  via radial kernels  $\phi_a$  and  $\phi_b$  (as in Eqs. 9 and 12), their Fourier- and harmonic-space power spectra are related through

$$\begin{aligned} \mathcal{S}_{\ell}^{ab} &= \frac{2}{\pi} \int dk k^2 \int d\chi_1 \phi_a(\chi_1) j_{\ell}(k\chi_1) \\ &\times \int d\chi_2 \phi_b(\chi_2) j_{\ell}(k\chi_2) P_{ab}(k; z_1, z_2), \end{aligned} \quad (\text{A.5})$$

where  $j_{\ell}$  is the spherical Bessel function of order  $\ell$ . For broad kernels, we can use the Limber approximation,  $j_{\ell}(x) \sim \sqrt{\pi/(2\ell+1)} \delta(\ell+1/2-x)$ , in which case the previous relation simplifies to Eq. (16).

## Appendix B: Optimal weights

Here we derive the choice of optimal weights discussed in Sect. 2.3.3. The derivation is a standard result in statistics and follows the discussion in Appendix A of Alonso et al. (2020).

We consider a vector of  $N$  measurements  $\mathbf{x} = (x_1, \dots, x_N)$  and the problem of finding the linear combination of this vector that provides the best estimator of a given quantity,  $y$ . Assuming Gaussian statistics, the conditional probability is

$$\log p(y|\mathbf{x}) = \mathbf{z}^T \mathbf{C}_{zz}^{-1} \mathbf{z} - \mathbf{x}^T \mathbf{C}_{xx}^{-1} \mathbf{x}, \quad (\text{B.1})$$

where  $\mathbf{z} \equiv (y, x_1, \dots, x_N)$ , and  $\mathbf{C}_{ab}$  is the covariance matrix between  $\mathbf{a}$  and  $\mathbf{b}$ <sup>4</sup>. The minimum-variance estimator for  $y$  given this distribution coincides with its mean, which is given by

$$\hat{y} = \mathbf{w}^T \mathbf{x} \equiv \mathbf{C}_{xy}^T \mathbf{C}_{xx}^{-1} \mathbf{x}. \quad (\text{B.2})$$

The linear coefficients  $\mathbf{w}$  are the so-called Wiener filter.

<sup>4</sup> We use boldface sans serif characters  $\mathbf{C}$  to denote matrices and boldface italic characters  $\mathbf{C}$  to denote vector quantities.

If  $y$  is the UHECR flux and  $\mathbf{x}$  is a set of galaxy over-density maps at different radial shells with co-moving width  $d\chi$ , in the shot noise-dominated regime  $\mathbf{C}_{xy}$  and  $\mathbf{C}_{xx}$  are given by

$$\mathbf{C}_{\text{CR},g}(\chi) \propto d\chi \frac{\alpha[z(\chi)]}{1+z(\chi)}, \quad (\text{B.3})$$

$$\mathbf{C}_{g,g}(\chi, \chi') \propto \delta_{\chi, \chi'} d\chi \chi^2 \bar{n}_{g,c}(\chi), \quad (\text{B.4})$$

where we have ignored all  $r$ -independent pre-factors. Therefore, the Wiener filter in this case is

$$w(\chi) = \frac{\alpha[z(\chi)]}{[1+z(\chi)] \chi^2 \bar{n}_{g,c}(\chi)}. \quad (\text{B.5})$$

## Appendix C: Halo occupation distribution

Halo occupation distribution models have been used profusely in the literature to describe the connection between the galaxy number density and the matter over-density. Here we give a brief overview of the specifics of the model used in this work to describe the low-redshift 2MRS sample, which follows Ando et al. (2018).

Within the halo model (Peacock & Smith 2000; Cooray & Sheth 2002), all matter in the Universe can be found in haloes of different masses, and therefore the fluctuations of a given quantity,  $x$ , can be described in terms of its distribution around haloes as a function of halo mass,  $u_x(r, M)$  (also called the halo profile), and the correlated distribution of haloes on large scales. In this formalism, the power spectrum between two quantities,  $x$  and  $y$ , receives two contributions, coming from the correlations between mass elements belonging to the same halo and mass elements in different haloes (the so-called one-halo and two-halo terms), as  $P_{xy} = P_{xy}^{\text{1h}} + P_{xy}^{\text{2h}}$ , with

$$P_{xy}^{\text{1h}}(k) = \int dM n_{\text{h}}(M) u_x(k, M) u_y(k, M), \quad (\text{C.1})$$

$$\begin{aligned} P_{xy}^{\text{2h}}(k) &= \left[ \int dM n_{\text{h}}(M) b_{\text{h}}(M) u_x(k, M) \right] \\ &\times \left[ \int dM n_{\text{h}}(M) b_{\text{h}}(M) u_y(k, M) \right] P_{\text{lin}}(k), \end{aligned} \quad (\text{C.2})$$

where  $n_{\text{h}}(M)$  and  $b_{\text{h}}(M)$  are respectively the halo mass function and the halo bias,  $P_{\text{lin}}(k)$  is the linear matter power spectrum, and  $u_x(k, M)$  is the Fourier transform of the halo profile.

In our case, we want to model the galaxy over-density,  $\delta_g$ , and therefore we need to specify the halo galaxy density profile. For this, we employed the formalism used in Ando et al. (2018):

$$u_g(k) = \frac{1}{\bar{n}_{g,c}} [N_c(M) + N_s(M) u_s(k, M)], \quad (\text{C.3})$$

where  $N_c(M)$  and  $N_s(M)$  are the number of central and satellite galaxies, the latter of which are distributed according to  $u_s(r)$ . Centrals and satellites are distributed according to Bernoulli and Poisson distributions, respectively. Their mean values and the satellite profile are modelled as a function of mass as:

$$\bar{N}_c(M) = \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{\log M - \log M_{\text{min}}}{\sigma_{\log M}} \right) \right], \quad (\text{C.4})$$

$$\bar{N}_s(M) = \left( \frac{M - M_{\text{min}}}{M_1} \right)^{\alpha} \Theta(M - M_{\text{min}}), \quad (\text{C.5})$$

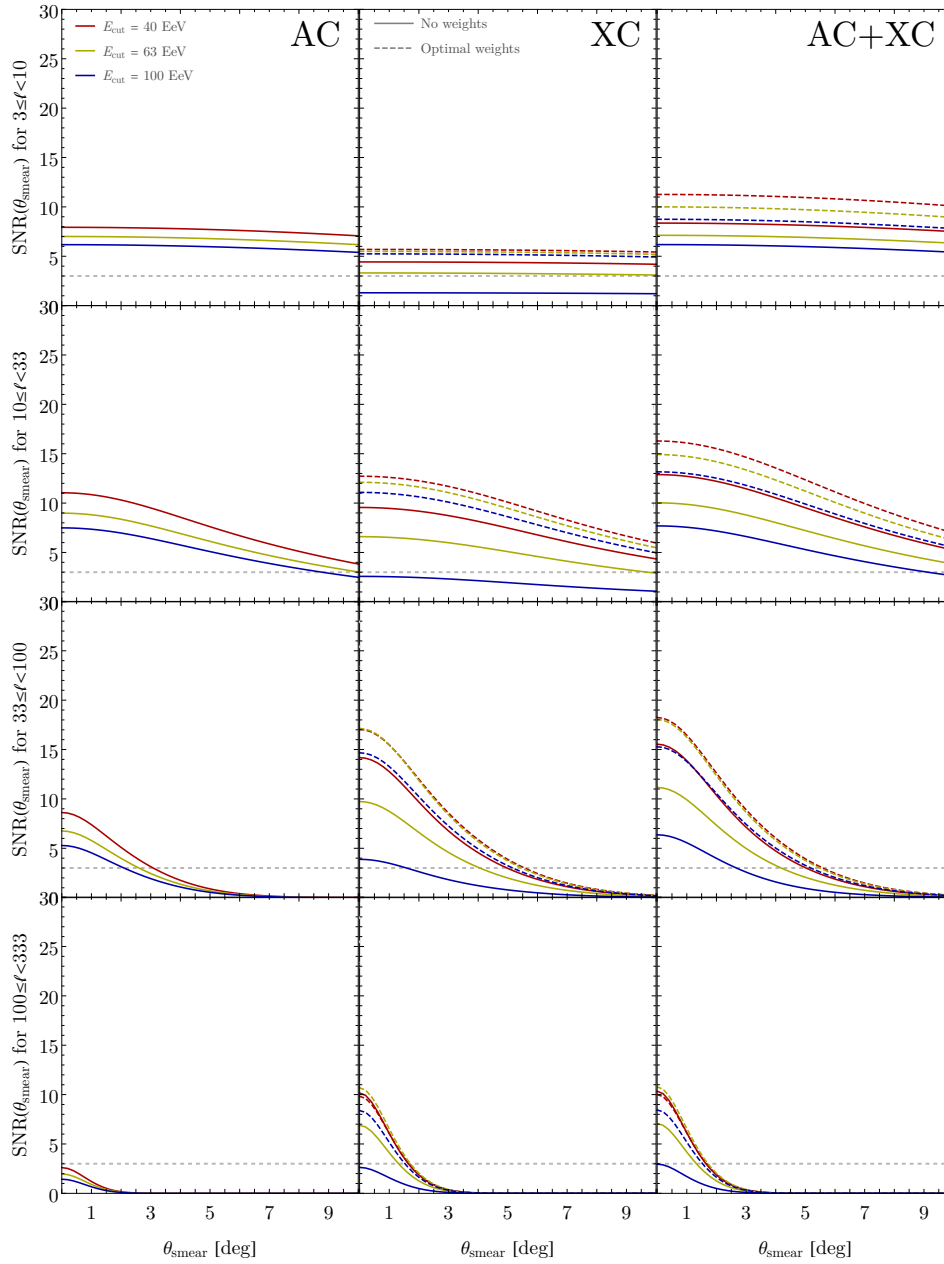
$$u_s(r, M) = \frac{\Theta(r_{\text{max},g} - r)}{(r/r_{s,g})(r/r_{s,g} + 1)^2}, \quad (\text{C.6})$$

where  $\Theta$  is the Heaviside function. The free parameters of the model are  $M_{\min}$ ,  $M_1$ ,  $\sigma_{\log M}$ ,  $r_{\max}$ ,  $g/r_s$ ,  $r_{s,g}/r_s$ , and  $\alpha$ , where  $r_s$  is the mass-dependent halo scale radius. We used the best-fit values found by Ando et al. (2018) for these parameters in our calculation.

## Appendix D: Magnetic deflections

The effects of the GMF deflections are not the same across harmonic multipoles. We expect that, for both the AC and the XC, small scales would be more affected by the deflections. We illustrate this in Fig. D.1, where we show the equivalent of Fig. 6 but

for multipoles in the four half decades:  $\ell \in [3, 10[$ ,  $\ell \in [10, 33[$ ,  $\ell \in [33, 100[$ , and  $\ell \in [100, 333[$ . As anticipated, the smearing suppresses the power at small scales more incisively, for both the AC and the XC, with the XC more affected. Nonetheless, in regions of the sky where the GMF is small, for example around the Galactic polar caps, the XC has better chances to be detected than the AC at large  $\ell$  (small scales). In reading this figure one should keep at least three simplifications in mind: The larger but physically different effects of the large-scale GMF are not included; the smearing angle is constant across the sky (see Eq. (26)); and the choice of a Gaussian smearing is arbitrary as other types of beams might reproduce the actual deflections more faithfully.



**Fig. D.1.** Total S/N per half-decade,  $\sqrt{\sum_{\ell=\ell_{\min}}^{\ell_{\max}} S/N_{\ell}^2}$ , with  $(\ell_{\min}, \ell_{\max})$  in  $[3, 10[$  (top row),  $[10, 33[$  (second row),  $[33, 100[$  (third row), and  $[100, 333[$  (bottom row), as a function of the smearing angle for the AC signal, the XC signal with both normal and optimal weights, and their combination, AC+XC (leftmost, central, and rightmost panel, respectively, in each row). Different colours refer to different energy cuts, and the horizontal dashed line shows the thresholds for  $3\sigma$  detection.