



**HAL**  
open science

# Retrieving cosmological information from small-scale CMB foregrounds - II. The kinetic Sunyaev Zel'dovich effect

Adélie Gorce, Marian Douspis, Laura Salvati

► **To cite this version:**

Adélie Gorce, Marian Douspis, Laura Salvati. Retrieving cosmological information from small-scale CMB foregrounds - II. The kinetic Sunyaev Zel'dovich effect. *Astronomy and Astrophysics - A&A*, 2022, 662, pp.A122. 10.1051/0004-6361/202243351 . hal-03594444

**HAL Id: hal-03594444**

**<https://hal.science/hal-03594444>**

Submitted on 24 Jul 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Retrieving cosmological information from small-scale CMB foregrounds

## II. The kinetic Sunyaev Zel'dovich effect

Adélie Gorce<sup>1</sup>, Marian Douspis<sup>2</sup>, and Laura Salvati<sup>2</sup>

<sup>1</sup> Department of Physics and McGill Space Institute, McGill University, Montreal, QC H3A 2T8, Canada  
e-mail: [adelie.gorce@mcgill.ca](mailto:adelie.gorce@mcgill.ca)

<sup>2</sup> Université Paris-Saclay, CNRS, Institut d'Astrophysique Spatiale, 91405 Orsay, France

Received 17 February 2022 / Accepted 22 April 2022

### ABSTRACT

Recent results of ground-based telescopes, giving high-quality measurements of the cosmic microwave background (CMB) temperature power spectrum on small scales (below 1 deg) motivate the need for an accurate model of foregrounds, which dominate the primary signal at these multipoles. In a previous work, we have shown that cosmological information could be retrieved from the power spectrum of the thermal Sunyaev Zel'dovich (SZ) effect. In this work, we introduce a physically motivated model of the Epoch of Reionisation in the cosmological analysis of CMB data, which is coherent on all scales. In particular, at high multipoles, the power spectrum of the kinetic SZ (kSZ) effect is inferred from a set of cosmological and reionisation parameters by a machine-learning algorithm. First including an asymmetric parameterisation of the reionisation history in the *Planck* 2018 data analysis, we retrieve a value of the Thomson optical depth consistent with previous results, but stemming from a completely different history of reionisation in which the first luminous sources light up as early as  $z = 15$ . Considering the latest small-scale data from the South Pole telescope (SPT) and letting the cosmology free to vary, we find that including the new cosmology-dependent tSZ and kSZ spectra helps tighten the constraints on their amplitudes by breaking their degeneracy. We report a  $5\sigma$  measurement of the kSZ signal at  $\ell = 3000$ ,  $\mathcal{D}_{3000}^{\text{kSZ}} = 3.4_{-0.3}^{+0.5} \mu\text{K}^2$  at the 68% confidence level (C.L.), marginalised over cosmology, as well as an upper limit on the patchy signal from reionisation  $\mathcal{D}_{3000}^{\text{pkSZ}} < 1.6 \mu\text{K}^2$  (95% C.L.). Additionally, we find that the SPT data favour slightly earlier reionisation scenarios than *Planck*, leading to  $\tau = 0.062_{-0.015}^{+0.012}$  and a reionisation midpoint  $z_{\text{re}} = 7.9_{-1.3}^{+1.1}$  (68% C.L.), which is in line with constraints from high-redshift quasars and galaxies.

**Key words.** methods: data analysis – methods: statistical – cosmic background radiation – dark ages, reionization, first stars

## 1. Introduction

Different generations of satellites, from the cosmic background explorer (COBE, [Bennett et al. 1996](#)) to, most recently, *Planck* ([Planck Collaboration I 2020](#)), led us to exquisite measurements of the power spectrum of the large-scale fluctuations of the cosmic microwave background (CMB) temperature. Today, efforts are turned towards ground-based observatories of such fluctuations on small scales. The Atacama cosmology telescope (ACT, [Kosowsky 2003](#))<sup>1</sup> and the South Pole telescope (SPT, [Ruhl et al. 2004](#))<sup>2</sup> target this signal, with the latter recently publishing observations of sufficient quality ([Reichardt et al. 2021](#), hereafter R21) to allow for the first  $\geq 3\sigma$  measurement of the amplitude of the angular power spectra of the kinetic and thermal Sunyaev–Zel'dovich effects (kSZ and tSZ, respectively, [Zeldovich & Sunyaev 1969](#); [Sunyaev & Zeldovich 1980](#)) at an angular multipole of  $\ell = 3000$ .

Along with thermal dust emission from dusty star-forming galaxies, radio galaxy emission, the Galactic cirrus signal, and the cosmic infrared background (CIB), the tSZ and kSZ power spectra are foregrounds to the CMB primary anisotropies. That

is, their power is not sourced by temperature fluctuations in the primordial plasma at the time of recombination, but from the interaction of the photons released at the surface of last scattering with astrophysical objects along the line of sight, before reaching the detector. In this respect, and despite the fact that they largely dominate the primary signal on scales smaller than one arcminute, their amplitudes have often been considered in analyses of CMB data as nuisance parameters (e.g., [Reichardt et al. 2012](#); [Dunkley et al. 2013](#); [Ichiki 2014](#); [George et al. 2015](#)). A common approach consists of fitting a primary power spectrum derived from sampled cosmological parameters and a sum of foregrounds power spectra modelled by templates to the observed data points. Only the amplitudes of the various foregrounds, at a given multipole (often,  $\ell = 3000$ ), are fitted to data.

However, these foregrounds trace the matter the CMB photons encounter on their way to us and they include cosmological information which cannot be retrieved by the use of templates ([Douspis et al. 2006](#); [Shaw et al. 2010](#); [Salvati et al. 2018](#); [Maniyar et al. 2019](#); [Gorce et al. 2020](#)). In a first work ([Douspis et al. 2022](#), hereafter Paper I), we have shown that replacing the template used for the tSZ power spectrum by a cosmology-dependent spectrum when analysing data from *Planck* and the SPT tightens constraints on cosmological

<sup>1</sup> <https://act.princeton.edu>.

<sup>2</sup> <http://pole.uchicago.edu>.

parameters. In this work, we perform a similar analysis, this time focusing on constraints obtained when including a cosmology-dependent kSZ angular power spectrum in the analysis.

The kSZ power spectrum traces the energy gained by CMB photons interacting with electrons that have a non-zero bulk velocity relative to the CMB rest-frame along the line-of-sight. Numerous works (e.g., McQuinn et al. 2005; Iliev et al. 2007; Mesinger et al. 2012; Shaw et al. 2012; Battaglia et al. 2013; Alvarez 2016; Chen et al. 2022) have shown that the amplitude of the kSZ angular power spectrum can be related to the nature of the Epoch of Reionisation (EoR), when the first light sources re-ionised the intergalactic medium (IGM), mostly neutral since recombination and the emission of the CMB photons. Indeed, as an integrated effect, a longer or earlier reionisation process will lead to a larger kSZ amplitude. Traditionally, the kSZ power is divided into two contributions: The late-time kSZ accounts for CMB photons scattering off electrons once reionisation is over and the IGM is fully ionised, whilst the patchy kSZ accounts for photons scattering off ionised bubbles along the otherwise neutral line of sight during reionisation. In contrast to the late-time kSZ, not only the amplitude, but also the shape of the patchy kSZ power spectrum can tell us about the reionisation process (McQuinn et al. 2005; Iliev et al. 2007; Mesinger et al. 2012). For example, the multipole at which the patchy kSZ reaches its maximum has been related to the characteristic size of ionised bubbles during the EoR, in turn related to the physical properties of early galaxies, if those are assumed to have led the reionisation process (Bouwens et al. 2015; Robertson et al. 2015; Greig et al. 2017; Gorce et al. 2018).

With the new generation of the SPT (Sobrin et al. 2022), the Simons Observatory (Ade et al. 2019) and, later, CMB-Stage 4 experiments (CMB-S4, Abazajian et al. 2016), coming online, there is a growing interest in the community for the potential of small-scale CMB data to constrain reionisation. A common approach, popularised by the sampler of 21cmFAST simulations (Mesinger et al. 2011; Park et al. 2019; Murray et al. 2020), 21CMMC (Greig & Mesinger 2015), is to sample astrophysical parameters in semi-numerical simulations and confront the observables derived at each iteration to data (Qin et al. 2020; Choudhury et al. 2021). However, such approaches are limited by computational power, both in terms of the precision and the size of the employed simulations. Very large box sizes are necessary to capture the large-scale velocity flows contributing to the kSZ effect at high- $\ell$  and results based on insufficiently large simulations will significantly underestimate its power (Shaw et al. 2012).

To circumvent these numerical difficulties, we have introduced in Gorce et al. (2020) a new analytical derivation of the kSZ angular power spectrum, based on the cosmology and the reionisation history. With this formalism, it is possible to compute the theoretical kSZ power spectrum associated with a set of cosmological and reionisation parameters at each iteration of the sampler, when analysing small-scale fluctuations of the CMB temperature. Constraints on the cosmological model and the EoR are retrieved from the measured shape and amplitude of the kSZ signal. Such an analysis, applied to the latest *Planck* and SPT data (Planck Collaboration I 2020; Reichardt et al. 2021), is the focus of the present work. In an approach similar to Paper I, we use machine learning to efficiently predict the otherwise computationally intensive derivation of the kSZ power spectrum on the multipole range covered by the data. In order for the reionisation modelling to be consistent across all scales probed, we substitute the commonly used ‘instantaneous’ parameterisation of the evolution of the IGM ionised fraction with redshift (Lewis

et al. 2000) by a physically motivated asymmetric parameterisation (Douspis et al. 2015), also used in the kSZ derivation. We subsequently assess the impact of this substitution on the cosmological constraints obtained with large-scale data.

This paper is organised as follows. In Sect. 2, we detail the theoretical framework used to derive the kSZ power spectrum, and in particular, introduce the asymmetric parameterisation of reionisation used throughout. We also introduce the statistical methods used to analyse the data listed in Sect. 3. In Sect. 4, we first present the results of applying the asymmetric reionisation model to *Planck* data. We then include the new kSZ modelling in the analysis of the small-scale SPT data and retrieve reliable constraints on the midpoint and duration of reionisation. Finally, we discuss our results in Sect. 5 and conclude in Sect. 6.

## 2. Methods

### 2.1. Asymmetric reionisation history

In the fiducial analysis of *Planck* large-scale data, the reionisation history is not constrained directly. Instead, it is the Thomson optical depth  $\tau$  that is sampled, as it is strongly correlated to the amplitude of the high- $\ell$  temperature and low- $\ell$  polarisation CMB power spectra (Planck Collaboration XLVII 2016). As a measure of the fraction of CMB photons absorbed by free electrons along the line-of-sight, the optical depth is directly related to the ionisation level of the IGM or number of electrons per hydrogen nucleus  $x_e(z)$  (Lewis et al. 2000; Howlett et al. 2012):

$$\tau(z) = c \sigma_T \int_0^z dz' x_e(z') n_b(z') \frac{(1+z')^2}{H(z')}, \quad (1)$$

where  $\sigma_T$  is the Thomson cross-section,  $n_b$  is the average density of baryons and  $H$  is the Hubble parameter. Here, for simplicity, the product of  $x_e$  and  $n_b$ , both averaged over the sky, is taken as a proxy for the mean electron density,  $\langle x_e \times n_b \rangle$ , overlooking the correlation between the electron and the density fields. This can lead to a 5–10% error on the measured value of  $\tau$  (Liu et al. 2016). To interpret  $\tau$  in terms of the reionisation history, the following symmetric, ‘instantaneous’, parameterisation is often considered:

$$x_e(z) = \frac{f_e}{2} \left[ 1 + \tanh \left( \frac{y - y_{\text{re}}}{\delta y} \right) \right], \quad (2)$$

where  $y(z) \equiv (1+z)^{\frac{3}{2}}$ ,  $y_{\text{re}} = y(z = z_{\text{re}})$  for  $z_{\text{re}}$  the midpoint of hydrogen reionisation, such that  $x_e(z_{\text{re}}) = 0.5$  and  $\delta y \equiv \frac{3}{2} (1+z)^{\frac{1}{2}} \delta z$  for  $\delta z = 0.5$ , roughly corresponding to the redshift range which sees the IGM ionised fraction increasing from 25% to 75%. The  $f_e$  parameter corresponds to the number of electrons per hydrogen atom, and is equal to  $f_e \simeq 1.08$  when accounting for the first ionisation of helium. The contribution of the second helium ionisation is added for  $z > 4.5$  as another hyperbolic tangent such as the one given in Eq. (2), with parameters  $z_{\text{re,He}} = 5.0$  and  $\delta z_{\text{He}} = 0.5$  (Kuhlen & Faucher-Giguère 2012). It is finished by  $z = 3.5$ . When  $\delta z$  is fixed, it is possible to deduce the value of  $z_{\text{re}}$  through binary search from a given  $\tau$  value.

In this framework, the evolution of  $x_e$  with redshift is a quick and symmetrical process, centred on  $z_{\text{re}}$ . However, recent simulations and analyses of astronomical data sets show the reionisation process to be a highly gradual evolution, starting slowly at redshifts as high as  $z = 15$  (Robertson et al. 2015; Gorce et al. 2018). Since the kSZ power depends largely on the duration of reionisation (see Sect. 2.2 and Battaglia et al. 2013;

Alvarez 2016) and on its asymmetry (Park et al. 2013; Chen et al. 2022), it is essential to include a more accurate evolution of  $x_e(z)$  in any analysis attempting to measure its amplitude. Accounting for this asymmetry can be done by, for instance, interpolating between different ionisation levels (e.g., Millea & Bouchet 2018; Trac et al. 2022) or by changing the parameterisation for the redshift-evolution of the ionisation level. In this work, we follow this second approach and use the power-law introduced in Douspis et al. (2015) and already confronted to large-scale CMB data in Planck Collaboration XLVII (2016):

$$x_e(z) = \begin{cases} f_e & \text{for } z < z_{\text{end}}, \\ f_e \left( \frac{z_{\text{early}} - z}{z_{\text{early}} - z_{\text{end}}} \right)^\alpha & \text{for } z > z_{\text{end}}, \end{cases} \quad (3)$$

where  $z_{\text{early}}$  corresponds to the redshift around which the first emitting sources form and at which  $x_e(z)$  is matched to a residual ionised fraction of  $10^{-4}$ . We fix  $z_{\text{early}} = 20$ , in agreement with current constraints on the history of reionisation obtained from astrophysical data sets (Robertson et al. 2015; Greig & Mesinger 2015; Gorce et al. 2018). Additionally, we find that the data considered in this work does not require sources to light up before  $z = 20$ . We let  $z_{\text{end}}$  and  $z_{\text{re}}$  vary, since fixing  $z_{\text{early}}$  makes  $z_{\text{re}}$  a sampling equivalent to  $\alpha$  (Eq. (3)), and we choose the former to facilitate the interpretation of results. Finally, for convergence purposes, we sample  $dz \equiv z_{\text{re}} - z_{\text{end}}$  instead of  $z_{\text{end}}$  and we consider the contribution of helium reionisation. We include this parameterisation into the derivation of the kSZ angular power spectrum detailed below.

## 2.2. Derivation of the kSZ power spectrum

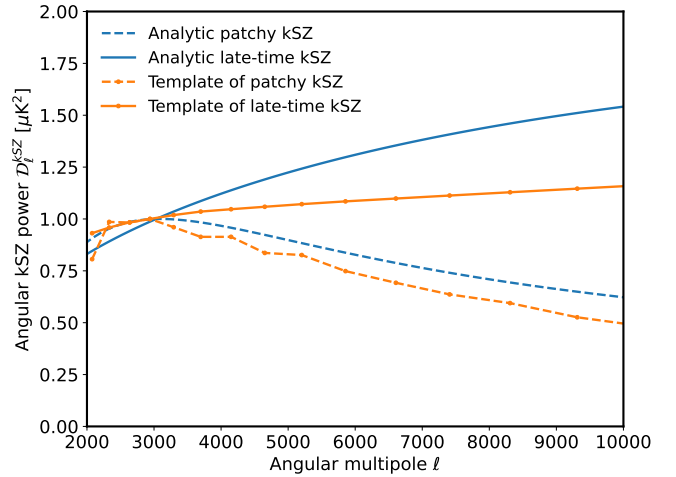
The temperature anisotropies sourced by the scattering of CMB photons off clouds of free electrons with a non-zero bulk velocity  $\mathbf{v}$  relative to the CMB rest-frame along the line of sight  $\hat{\mathbf{n}}$  write

$$\delta T_{\text{kSZ}}(\hat{\mathbf{n}}) = \frac{\sigma_T}{c} \int \frac{d\eta}{dz} \frac{dz}{(1+z)} e^{-\tau(z)} n_e(z) \mathbf{v} \cdot \hat{\mathbf{n}}, \quad (4)$$

with  $\eta$  the comoving distance to redshift  $z$  and  $\mathbf{v} \cdot \hat{\mathbf{n}}$  the component of the peculiar velocity of the electrons along the line of sight. The corresponding angular power spectrum writes  $C_\ell^{\text{kSZ}} \equiv T_{\text{CMB}}^2 |\delta \hat{T}_{\text{kSZ}}(k)|^2$ , where  $T_{\text{CMB}}$  is the mean temperature of the CMB and  $\delta \hat{T}_{\text{kSZ}}$  is the Fourier transform of  $\delta T_{\text{kSZ}}$  given above, such that  $k = \ell/\eta$  for  $\ell$  the angular multipole. In the following, we present results in terms of  $\mathcal{D}_\ell \equiv \ell(\ell+1)C_\ell/2\pi$ .

As already mentioned in Sect. 1, the kSZ signal is often decomposed into a late-time and a patchy component. In Gorce et al. (2020), we introduced an analytical derivation of the kSZ angular power spectrum based on the reionisation history  $x_e(z)$  and two parameters related to the morphology of reionisation. Firstly, the drop-off frequency of the power spectrum of free electrons density fluctuations,  $\kappa$ , which is directly related to the angular scale where the patchy kSZ power reaches its maximum, and, in turn, to the typical size of ionised bubbles during the EoR. It is expressed in  $\text{Mpc}^{-1}$ . Secondly, the large-scale amplitude of the electron power spectrum,  $\alpha_0$ , which is a representation of the variance of the electron field and, as such, is related to the morphology of the ionisation field. It is expressed in  $\text{Mpc}^3$ .

These two parameters are directly related to the shape and amplitude of the patchy component of the kSZ spectrum. The late-time contribution is derived using a biased matter power spectrum to describe the low-redshift spatial distribution of free electrons, according to the model introduced in Shaw et al.



**Fig. 1.** Comparison of the normalised kSZ angular power spectra obtained for a set of reionisation parameters using binned templates (Reichardt et al. 2021), in orange, and using the analytic derivation introduced in Gorce et al. (2020), in blue.

(2012), fitted to highly resolved simulations (Aubert et al. 2015) – see Gorce et al. (2020) for more details on the model. Instead of using a template of the kSZ power spectrum and fitting its amplitude to small-scale CMB data, in this work, we sample  $\alpha_0$  and  $\kappa$ , along with  $z_{\text{re}}$  and  $z_{\text{end}}$ , so that the kSZ spectrum can be fully computed at each iteration of the sampler. In general, we expect the modelling uncertainties of the parameterisation, including the additional kSZ power stemming from the non-Gaussianity of the electron density field (Alvarez 2016) to be absorbed by the uncertainties on the sampled parameters. However, once observations and foreground-modelling allow for a detailed detection of the kSZ spectra, a more precise description will be required.

In Fig. 1, we show the difference in the late-time and patchy kSZ spectra corresponding to the templates used in R21, and obtained using the analytic derivation of Gorce et al. (2020) for parameters fitted on a semi-numerical simulation (Seiler et al. 2019). In order to highlight shape differences, the spectra are normalised to  $1 \mu\text{K}^2$  at  $\ell = 3000$ . The template used for the late-time signal is extracted from the work of Shaw et al. (2012), who fit the power spectrum of electron number density fluctuations, considered as a biased matter power spectrum, on hydrodynamic simulations. The template used for the patchy signal was first introduced in Zahn et al. (2012) and was computed numerically from a full radiative transfer simulation (Zahn et al. 2007, 2011). There is a clear difference in the shapes obtained in both cases, in particular on small scales. This is related to the fact that the low-redshift section of our analytic kSZ model was calibrated on a simulation which uses adaptive mesh refinement, and, hence, resolves high-density regions very well, making the derived spectra more sensitive to the thermal behaviour of the gas, and boosting power on small scales (Aubert et al. 2015). The results presented in Sect. 4 are, however, robust to these modelling uncertainties.

The cosmology-dependent derivation of the kSZ power spectrum is computationally expensive because of the complexity of Eq. (4), which includes several integrals. Additionally, the time required to compute a  $C_\ell$  increases with  $\ell$ , making the analysis of small-scale data challenging. For this reason, it cannot be included as such in a parameter sampling algorithm. Similarly to what was proposed in Paper I, we choose to solve this issue



by using a machine learning algorithm, random forests (RF), trained on a set of kSZ spectra computed according to Gorce et al. (2020), to predict the  $C_\ell$ 's at each iteration of the sampler. The training set is made of 50 000 different spectra, computed on a range of cosmological and reionisation parameters, for 25 angular multipoles linearly distributed over the range  $100 \leq \ell \leq 10\,500$ , which includes the SPT data points. The details of the RF training and testing can be found in Appendix A. Namely, the errors coming from the RF reconstruction never exceed  $0.02 \mu\text{K}^2$ , about fifty times smaller than the precision of current measurements (Reichardt et al. 2021). The overall scores of the patchy and late-time predictions are, respectively, 99.3% and 99.8%. The RF prediction code for the kSZ power spectra is publicly available<sup>3</sup>.

### 2.3. MCMC analysis

We use a version of the publicly available Monte-Carlo Markov chain sampler CosmoMC (Lewis & Bridle 2002; Lewis 2013), modified to include the SPT likelihood, available online (Reichardt et al. 2021), and the *Planck*21-tSZ likelihoods, based on Tanimura et al. (2022) data and following Salvati et al. (2018). We further modify the sampler to allow for an on-the-fly RF-prediction of the tSZ, patchy kSZ and late-time kSZ angular power spectra (see above and Appendix A), as initiated in Paper I. For the derivation of the tSZ power spectrum, we follow the recipe detailed in Paper I. We add in the model the scaling relation between the tSZ signal (the Compton parameter  $Y_{500}$ ) and the halo mass  $M_{500}$ , both defined within a radius that contains 500 times the critical density of the Universe. In this analysis, for the scaling relation we include as a varying parameter only the mass bias  $(1 - b)$ , defined as the ratio between the halo mass estimated from the hydrostatic equilibrium assumption  $M_{\text{HE}}$ , and the total halo mass, that is  $(1 - b) = M_{\text{HE}}/M_{500}$ . We focus on this parameter, since it shows the strongest degeneracy with the cosmological ones. We refer the reader to the full discussion in Paper I. For the derivation of the kSZ power spectrum, we add as varying parameters in the analysis the two quantities introduced in Sect. 2.2, that is  $\log \alpha_0$  and  $\kappa$ . These replace the amplitude of the total kSZ power at  $\ell = 3000$ ,  $\mathcal{D}_{3000}^{\text{kSZ}}$ , which is now derived.

We modify the CAMB module (Lewis et al. 2000; Howlett et al. 2012), included within CosmoMC to perform theoretical computations, to account for the asymmetric parameterisation of the reionisation history given in Eq. (3). Doing so, the optical depth  $\tau$  is no longer sampled, but derived from the values of the reionisation mid- and endpoint,  $z_{\text{re}}$  and  $z_{\text{end}}$ , themselves sampled. We impose a prior  $z_{\text{end}} > 4.5$ . When the cosmology is let free to vary, we additionally sample the following parameters:

- The baryon density multiplied by the reduced Hubble constant  $h \equiv H_0/100$ ,  $\Omega_b h^2$ ;
- The cold dark matter density  $\Omega_c h^2$ ;
- The ratio of the sound horizon to the angular diameter distance at decoupling, scaled by 100,  $100\theta_{\text{MC}}$ ;
- The initial super-horizon amplitude of curvature perturbations at  $k = 0.05 \text{ Mpc}^{-1}$ ,  $A_s$ , as  $\log A \equiv \ln(10^{10} A_s)$ , and
- The scalar spectral index  $n_s$ .

Finally, nuisance parameters are added depending on the data set used, as described below.

The convergence of the runs is assessed by examining the Gelman–Rubin convergence parameter  $R$  (Gelman & Rubin

1992; An et al. 1998), which corresponds to the ratio of the variance of chain means to the mean of chain variances, for each sampled parameter. We consider the chains converged once, for each parameter,  $R - 1 < 10^{-2}$ . Unless specified otherwise, all parameter values are given in terms of the maximum of the marginalised posterior probability, and confidence intervals correspond to the highest-density probability intervals at 68%<sup>4</sup>.

## 3. Data

### 3.1. Planck low- $\ell$ data

For the analysis of large-scale data presented in Sect. 4.1, we consider the *Planck* Collaboration I (2020) data, available in the *Planck* Legacy archive<sup>5</sup>. We use the low- $\ell$   $TT$  and  $EE$  (small) likelihoods, as well as high- $\ell$   $TT$ ,  $TE$  and  $EE$  spectra (plik) and lensing. The cosmological constraints resulting from the fiducial analysis performed by the *Planck* collaboration are the following (with a 68% confidence):  $\Omega_b h^2 = 0.02237 \pm 0.00015$ ,  $\Omega_c h^2 = 0.1200 \pm 0.0012$ ,  $100\theta_{\text{MC}} = 1.04092 \pm 0.00031$ ,  $\tau = 0.0544 \pm 0.0073$ ,  $\log A = 3.044 \pm 0.014$  and  $n_s = 0.9649 \pm 0.0042$  (Planck Collaboration VI 2020).

### 3.2. SPT high- $\ell$ data

We consider the small-scale measurements of the CMB temperature power spectrum presented by the South Pole telescope collaboration in Reichardt et al. (2021) and available online<sup>6</sup>. The data are made of bandpowers at 95, 150 and 220 GHz measured over the angular multipole range  $2000 < \ell < 13\,500$ . The likelihood makes use of auto- and cross-spectra and marginalises over calibration and beam parameters. The total measured temperature power spectrum is considered to be the sum of the primary signal, as well as the following foregrounds: the thermal and kinetic Sunyaev–Zel’dovich effects, Poisson-distributed radio galaxies, whose power goes as  $\ell^2$ , and the galactic cirrus power. The contribution from thermal dust emission coming from the faint dusty star forming galaxies that cannot be masked out, and forming the CIB (Lagache et al. 2005; Casey et al. 2014), is also considered. The power is divided between spatially clustered and Poisson-distributed galaxies (Viero et al. 2013), with the latter following  $\mathcal{D}_\ell \propto \ell^2$ . These foregrounds, in the initial analysis, are all modelled with templates, whose amplitude at  $\ell = 3000$  is adjusted to fit the data. To this list is added the (negative) cross-spectrum of the CIB and the tSZ signal (Addison et al. 2012) and derived as a combination of the Shang et al. (2012) model and the tSZ  $C_\ell$ 's. The absolute value of the amplitude at  $\ell = 3000$  of the cross-spectrum, labelled  $\xi$ , is sampled. This results in a list of eight nuisance parameters (Fig. A.3).

## 4. Results

### 4.1. Substituting the reionisation model in low- $\ell$ data

First, we study the impact of changing the way reionisation is modelled when analysing the *Planck* Collaboration I (2020) data

<sup>4</sup> In contrast to equal-tail probability intervals, the highest-density probability interval is more appropriate for Bayesian analysis with skewed distributions, as it ensures all the values included in the confidence interval have a probability higher than 0.68.

<sup>5</sup> See <http://pla.esac.esa.int/pla/>.

<sup>6</sup> Bandpowers and likelihood codes compatible with CosmoMC can be downloaded at <https://pole.uchicago.edu/public/data/reichardt20/>.

<sup>3</sup> See <https://szdb.osups.universite-paris-saclay.fr/> for a user-friendly interface and a link to the public repository.

**Table 1.** Maximum likelihood parameters obtained with *Planck* CMB data when an asymmetric parameterisation of reionisation is used (see text).

Parameter	ML value
$\Omega_b h^2$	$0.0224 \pm 0.0001$
$\Omega_c h^2$	$0.1199 \pm 0.0012$
$100\theta_{MC}$	$1.0409 \pm 0.0003$
$n_s$	$0.9655 \pm 0.0040$
$\ln(10^{10} A_s)$	$3.048 \pm 0.014$
$z_{re}$	$7.4 \pm 0.7$
$dz$	$1.4^{+0.7}_{-0.9}$
$\tau$	$0.056 \pm 0.007$
$z_{end}$	$5.8^{+0.6}_{-1.2}$

**Notes.** The lower part of the table corresponds to derived parameters. Error bars are 68% confidence intervals.

described in Sect. 3.1. We compare the results obtained using the common instantaneous parameterisation of the reionisation history given in Eq. (2), on one side, and using the physically motivated asymmetric parameterisation given in Eq. (3), on the other side. We sample the following cosmological parameters:  $\log A$ ,  $\Omega_b h^2$ ,  $\Omega_c h^2$ ,  $100 \theta_{MC}$  and  $n_s$ . Additionally, we sample the parameters related to the history of reionisation, given in Eq. (3), that is  $z_{re}$  and  $z_{end}$ , with a prior  $z_{end} > 4.5$ . Here,  $\tau$  is a derived parameter since the reionisation history is directly probed. Templates are used for all the foregrounds.

The recovered maximum likelihood parameter values, along with their 68% confidence intervals, are given in Table 1. We find that changing the reionisation model has no impact on the baryon and dark matter densities, as well as on  $100 \theta_{MC}$ . As observed in previous works (e.g., Planck Collaboration XLVII 2016; Planck Collaboration VI 2020), we exactly recover the cosmological constraints obtained in Planck Collaboration I (2020) with the  $\tanh$  model, that is  $\Omega_b h^2 = 0.0224 \pm 0.0001$ ,  $\Omega_c h^2 = 0.1199 \pm 0.0012$  and  $100 \theta_{MC} = 1.0409 \pm 0.0003$  at  $1\sigma$ . Figure 2 compares the posterior distributions of the other four cosmological parameters sampled, for the two analyses. We see that the correlation between  $\tau$  and  $A_s$ , commonly observed in CMB analyses, is here translated into a correlation with  $z_{re}$ . Overall, results are similar for the two cases, since the low- $\ell$  CMB data are only really sensitive to the value of  $\tau$  and not to the details of the reionisation history, a point we further discuss in Sect. 5. Therefore, the errors on the optical depth are unchanged, as already observed by Planck Collaboration XLVII (2016) with *Planck* 2015 data. We present the most likely reionisation history, as well as the redshift-evolution  $\tau(z)$ , for each case, in the right panels of Fig. 2. In the asymmetric case (Eq. (3)), the extended tail towards high redshifts is compensated by a later reionisation, and a lower  $z_{re}$ , in order to reach an integrated  $\tau$  value similar to the one obtained with the  $\tanh$  model. This results in a completely different picture of the early Universe, with the first light sources starting reionising the IGM as early as  $z = 13$ . The first half of the reionisation history is well constrained, with  $z_{re} = 7.4 \pm 0.7$ , whilst there is a lot of uncertainty on the end of the process:  $z_{end} = 5.8^{+0.6}_{-1.2}$ . However, we can place an upper limit of  $z_{end} < 7.5$  at the 95% confidence level. These results show that a clear picture of reionisation cannot be obtained with large-scale CMB data only, but must be complemented by the cosmological information enclosed in small-scale data and foregrounds.

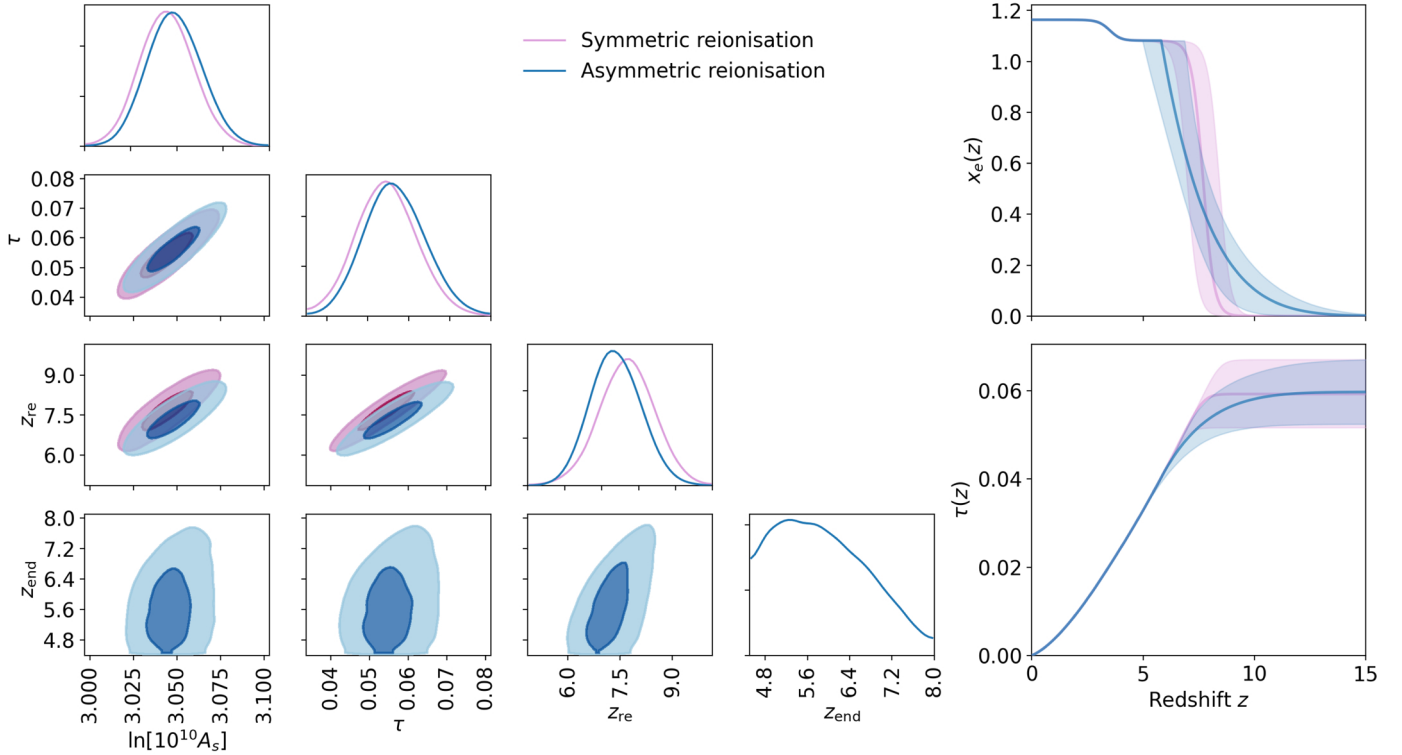
## 4.2. Substituting the kSZ template in high- $\ell$ data

### 4.2.1. Fixed cosmology

We now consider measurements of the CMB temperature fluctuations power spectrum on small scales obtained by the SPT and change the way the kSZ and tSZ signals are modelled in the analysis. Their angular power spectra are estimated using random forests (RF) trained on  $C_\ell$  values computed numerically from a set of cosmological, cluster physics and reionisation parameters (see Paper I, Sect. 2.2 and Appendix A): The kSZ shape parameters,  $\kappa$  and  $\log \alpha_0$ , as well as the mass bias  $1 - b$ , are now let free to vary on a range where the random forests have been trained. All the sampled parameters, including  $z_{re}$  and  $dz$ , are listed in Table 3. In order to compare with the results of Reichardt et al. (2021), we fix the cosmology to the values used by the authors:  $\Omega_c h^2 = 0.1193$ ,  $\Omega_b h^2 = 0.0224$ ,  $n_s = 0.9665$ ,  $\log A = 3.047$  and  $100\theta_{MC} = 1.04101$ . Here, it is not possible to fix the value of the Thomson optical depth as done in R21 since, as described in the previous paragraph, the reionisation model has been altered to make  $\tau$  a derived parameter. Instead, we impose a narrow Gaussian prior  $\tau = 0.056 \pm 0.007$ , the width of which corresponds to the errors on  $\tau$  coming from large-scale data (see Sect. 4.1).

We compare three cases. First, we replicate the results of R21, fixing cosmology and, effectively, reionisation. We then change the reionisation model for the one in Eq. (3) but keep templates for the tSZ and kSZ spectra and sample their amplitudes at  $\ell = 3000$ . Finally, we replace the templates by the RF-inferred tSZ and kSZ spectra, predicted at each iteration of the sampler.

In the top panel of Fig. 3, we show the resulting posterior distributions of the amplitudes of the tSZ and total kSZ power spectra at  $\ell = 3000$  and 143 GHz, labelled as  $\mathcal{D}_{3000}^{tSZ}$  and  $\mathcal{D}_{3000}^{kSZ}$ , respectively. Maximum likelihood values, along with 68% credibility intervals, are given in Table 2. Similarly to what was observed in Paper I, including the cosmology dependence of the SZ spectra in the analysis limits the values their amplitudes can take and breaks their degeneracy, leading to tighter constraints. We move from a  $3\sigma$  measurement reported in R21 to a  $6\sigma$  measurement of the amplitude of the kSZ signal and the error bars are cut in half:  $\mathcal{D}_{3000}^{kSZ} = 3.5^{+0.4}_{-0.2} \mu K^2$ . The RF prediction of the kSZ signal allows to decompose it between the contribution of post-reionisation (late) and patchy signals. In the bottom panel of Fig. 3, we present the probability distributions of their amplitude at  $\ell = 3000$ . The former largely dominates the total kSZ power, contributing to about 86% at  $\ell = 3000$  with  $\mathcal{D}_{3000}^{kSZ} = 3.1^{+0.1}_{-0.3} \mu K^2$ . Because of its low amplitude, only an upper limit can be placed on the latter in this setting:  $\mathcal{D}_{3000}^{pSZ} < 1.5 \mu K^2$  at 95% confidence. Regarding the tSZ amplitude, the uncertainty is reduced by 40% compared to R21, and we move from a 6 to a  $10\sigma$  measurement, such that  $\mathcal{D}_{3000}^{tSZ} = 3.3 \pm 0.3 \mu K^2$ . The distributions of other foreground parameters, shown in Fig. A.3, are also impacted by the change in the tSZ and kSZ modelling. For example, the amplitude of the contribution of dusty galaxies to the observed temperature power spectrum is better constrained, whilst staying consistent with previous results. The constraints on the amplitude of the tSZ–CIB correlations do not change whether we use templates or not ( $\xi = 0.07 \pm 0.03$ ), although the degeneracy of this parameter with  $\mathcal{D}_{3000}^{kSZ}$  and  $\mathcal{D}_{3000}^{tSZ}$  is lifted when templates are replaced by RF predictions. In agreement with R21, the data favour positive values of  $\xi$ , and a negative tSZ–CIB correlation, corresponding to an overdensity of dusty galaxies in galaxy clusters.



**Fig. 2.** Results obtained when fitting the large-scale *Planck* data with two different reionisation models: a physically motivated asymmetric parameterisation of  $x_e(z)$  (blue contours, Eq. (3)), or an instantaneous reionisation history (pink contours, Eq. (2)). *Left panel:* posterior distributions of cosmological parameters. *Right panels:* resulting redshift-evolution of the IGM ionised fraction (*top*) and optical depth (*bottom*), with associated 68% confidence intervals.

The cosmological information included in the high- $\ell$  SPT data is sufficient to give good constraints on the midpoint of reionisation and we find that, for a fixed cosmology, the use of a full derivation of the SZ power spectra instead of templates does not help constraining any of the reionisation parameters better. That is because changing the model but fixing the cosmology is equivalent to using another template, which R21 has shown to have no impact on the results. The model favoured by the small-scale data in this setting is one of early reionisation ( $z_{\text{re}} = 8.5^{+0.5}_{-0.4}$ ), in tension with currently available data points on the ionisation fraction of the IGM (see Fig. 8) and can be traced back to large optical depth values ( $\tau = 0.069 \pm 0.004$ ), over  $3\sigma$  away from the centre of the Gaussian prior imposed to the fit, and from the *Planck Collaboration I* (2020) constraints. This is due to the fact that on the  $\ell$ -range covered by the SPT data, there is a strong correlation between the  $A_s$  and  $\tau$  parameters, as the amplitude of the primary temperature power spectrum follows  $A_s \exp(-2\tau)$ . Fixing  $A_s$  is therefore equivalent to fixing  $\tau$  and the chosen value of  $A_s$  corresponds to a large value of  $\tau$ , which in turn triggers early reionisation scenarios and large kSZ amplitudes. We confirm this intuition by running the MCMC with RF-predicted tSZ and kSZ spectra and fixed cosmology, only changing the value of  $A_s$  to a lower value. This set-up leads to lower values of  $\tau$ , and, in turn, of the reionisation midpoint and of the patchy kSZ amplitude.

#### 4.2.2. Free cosmology

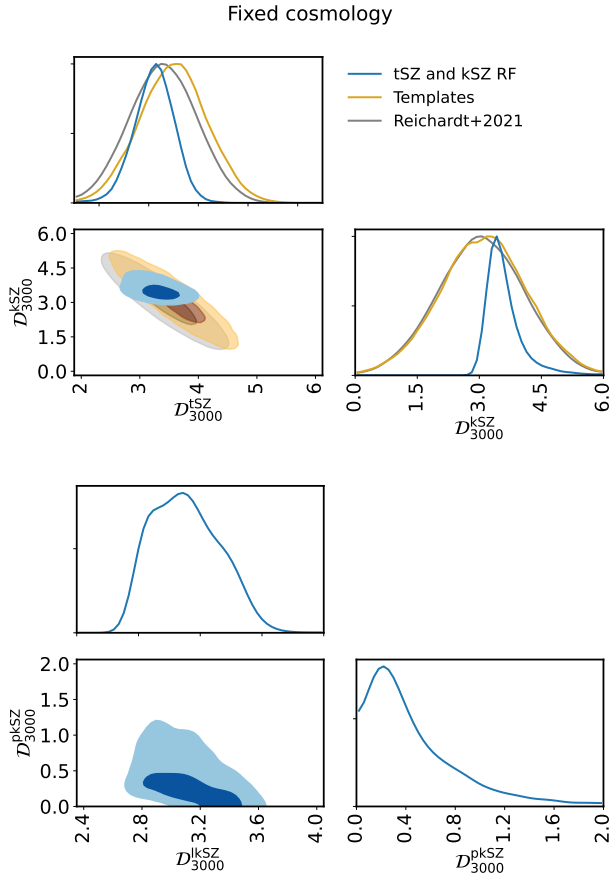
From the results presented in previous section, it is clear that the constraints on SZ amplitudes and on reionisation obtained with the SPT data strongly depend on the assumed cosmological model. Therefore, we now let cosmological parameters free but

apply tight Gaussian priors corresponding to the posterior distributions of *Planck Collaboration I* (2020) on  $\Omega_b h^2$ ,  $\Omega_c h^2$  and  $\theta_{\text{MC}}$ , which we have found in Sect. 4.1 to be largely impervious to changes in the reionisation history. On the other hand,  $\log A$  is left completely free, along with the reionisation parameters and the mass bias  $1 - b$ , related to the computation of the tSZ spectrum<sup>7</sup>. The applied priors are listed in Table 3. To assess the constraining power of our newly introduced cosmology-dependent tSZ and kSZ power spectra, we compare several cases, where either: Templates are used for both SZ spectra, a template is used for one SZ effect, whilst the other is predicted by the RF, or both SZ spectra are predicted by the RF. The latter set-up leads to the main results of this paper. The full list of maximum likelihood parameters resulting from these four analyses is given in Table 3 and a complete triangle plot of the posterior distributions of all parameters sampled, apart from nuisance parameters, can be found in Fig. B.1.

In the upper panel of Fig. 4, we show the posterior probability distributions of the tSZ and total kSZ amplitudes at  $\ell = 3000$  for the four considered cases. Again, replacing one of the two SZ templates by a cosmology-dependent spectrum is sufficient to break the degeneracy between the two amplitudes, as they can only take values compatible with the cosmology. However, here, replacing the tSZ template but keeping the kSZ one does not improve detection limits, with  $\mathcal{D}_{3000}^{\text{kSZ}}$  remaining at  $2\sigma$ . This can be related to the strong priors placed on cosmological parameters, in particular on  $\Omega_c h^2$  and, therefore, on  $\Omega_m$ , almost fixing the shape of the tSZ spectrum. Because of the strong

<sup>7</sup> As shown in Paper I, it is useful to keep in mind that the bias  $1 - b$  is strongly correlated to the amplitude of the tSZ power spectrum at  $\ell = 3000$  (see also Salvati et al. 2019).





**Fig. 3.** Constraints on the amplitude of the tSZ and kSZ power spectra at  $\ell = 3000$ , in  $\mu\text{K}^2$ , obtained with SPT high- $\ell$  data (R21) using either templates (in yellow) or RF-inferred tSZ and kSZ power spectra (in blue), as well as a physical parameterisation of reionisation. Inferences are compared to the initial analysis, with templates and fixed reionisation (in grey). For all cases, the cosmology is fixed (see text for values). On the *bottom panel*, the kSZ is decomposed into its late-time and patchy components.

correlation existing between  $\Omega_c h^2$  and the mass bias  $1-b$ , releasing the prior on  $\Omega_c h^2$  will weaken the constraints obtained on the mass bias, but have no impact on the measured  $\mathcal{D}_{3000}^{\text{tSZ}}$ . This also explains why the results only slightly change when the tSZ model changes but RF-inferred kSZ is used.

It is only when the cosmological information included in the kSZ spectrum is used that there is a clear improvement in the measurement of the two signals, with the most likely values of the tSZ and kSZ amplitudes sitting at 9 and 5 $\sigma$ , respectively, from a zero amplitude:  $\mathcal{D}_{3000}^{\text{tSZ}} = 3.5 \pm 0.4$  and  $\mathcal{D}_{3000}^{\text{kSZ}} = 3.4^{+0.5}_{-0.3}$  with a 68% confidence limit (see also Table 3). The model used to derive the kSZ allows to decompose it into its patchy and late-time contributions. The probability distributions of their amplitudes at  $\ell = 3000$  are shown in the bottom panel of Fig. 4. Again, if the late-time contribution is well-constrained, only an upper limit can be placed on the patchy kSZ, similar to what was obtained with a fixed cosmology:  $\mathcal{D}_{3000}^{\text{pkSZ}} < 1.6 \mu\text{K}^2$  at 95% confidence. The most likely kSZ and tSZ spectra at 150 GHz, as well as their 68% credibility regions, are shown in Fig. 5 and compared to the best-fit templates. For the first time, it is possible to see the shape of the SZ spectra preferred by the small-scale CMB data, and to decompose the kSZ power into its late-time and patchy components. In conclusion, in comparison to the templates, the RF-prediction allows for a larger power at high- $\ell$  for

**Table 2.** Most likely values of the amplitude, in  $\mu\text{K}^2$ , of the SZ effects obtained when fitting the SPT data.

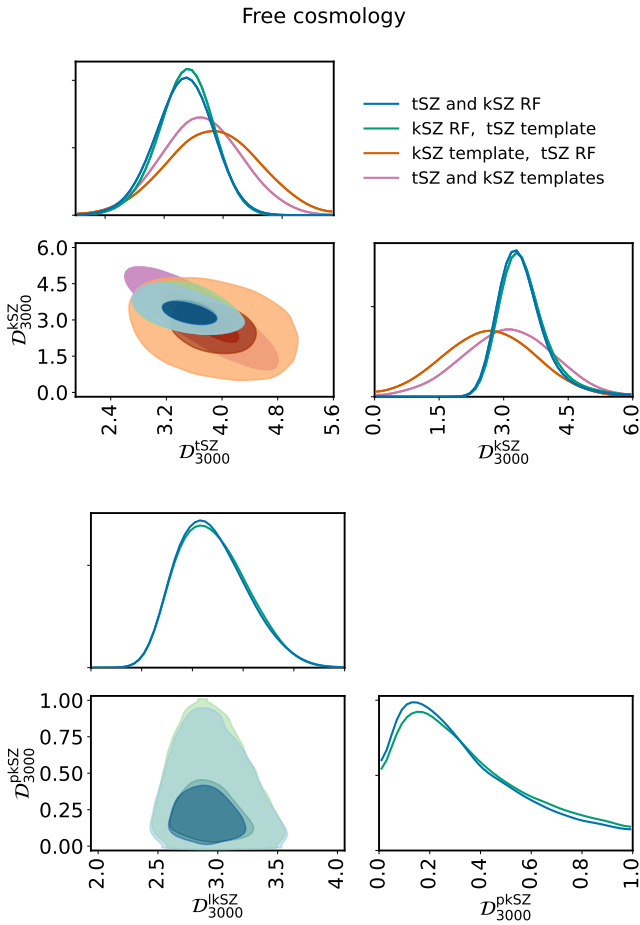
	R2021	Temp.	RF
$\mathcal{D}_{3000}^{\text{tSZ}}$	$3.4 \pm 0.5$	$3.6 \pm 0.5$	$3.3 \pm 0.3$
$\mathcal{D}_{3000}^{\text{kSZ}}$	$3.1 \pm 1.0$	$3.1^{+1.0}_{-1.1}$	$3.5^{+0.4}_{-0.2}$
Patchy	–	–	$< 1.5$
Late-time	–	–	$3.1^{+0.1}_{-0.3}$

**Notes.** Different cases are compared: the initial analysis by Reichardt et al. (2021), augmented by the asymmetric reionisation model (‘Temp.’), both using templates for the tSZ and the kSZ. ‘RF’ corresponds to the use of RF-inferred, cosmology-dependent tSZ and kSZ spectra. In all cases, the cosmology is fixed.

both spectra and for reduced error bars at all multipoles. This result holds when using different templates than the ones used in Reichardt et al. (2021), notably templates obtained with the analytical derivation of Gorce et al. (2020) for different sets of parameters. We additionally find a difference of only  $0.17 \mu\text{K}^2$ , corresponding to approximately  $0.25\sigma$ , on the amplitude of the total kSZ power at  $\ell = 3000$  when a template and a cosmology-dependent amplitude are used for the late-time component of the kSZ (Shaw et al. 2012), whilst the RF is used to infer the patchy contribution.

We show in Fig. 6 the maximum likelihood components of the temperature power spectrum observed by the SPT (R21) for our main analysis, that is when both SZ spectra are predicted by the RF, and when templates are used. We see that the SZ amplitudes at  $\ell = 3000$  are consistent across all analyses, but that freeing the shape of the SZ spectra when introducing the RF leads to an increase in the contribution of both the kSZ and the tSZ to the total power on the smallest scales ( $\ell > 4000$ ). This power is mostly drawn from clustered dusty galaxies and radio galaxies. The fixed shape of their spectra results in decreased amplitudes at  $\ell = 3000$ , with the amplitude of the contribution from radio galaxies moving from  $\mathcal{D}_{3000}^{\text{rg}} = 0.9 \pm 0.2$  to  $\mathcal{D}_{3000}^{\text{rg}} = 0.7 \pm 0.1$  when introducing the RF. The amplitudes of the other foregrounds are only weakly impacted by the new model. For example, the amplitude of the contribution from the Galactic cirrus at  $\ell = 3000$  and 220 GHz is  $\mathcal{D}_{3000}^{\text{gc}} = 1.9 \pm 0.4 \mu\text{K}^2$ . Only the constraints on the CIB-tSZ correlation, whose amplitude is described by a sampled parameter  $\xi$  (see Sect. 3), are strongly impacted by the change of model. We show in Fig. B.2 the joint posterior distributions of this parameter with the kSZ and tSZ amplitudes at  $\ell = 3000$  in the four cases aforementioned. In the analysis of previous data, allowing for a correlation between the two signals has made measuring the kSZ signal more difficult (Reichardt et al. 2012; Dunkley et al. 2013; Sievers et al. 2013; George et al. 2015), until R21, where the authors claim to measure the kSZ amplitude at  $3\sigma$ , despite letting  $\xi$  free. The previous difficulties to measure the SZ amplitudes are explained by a strong degeneracy between  $\xi$  and the kSZ and tSZ power. Here, when using both templates, we indeed find a correlation coefficient of 0.65 and  $-0.58$  between  $\xi$  and  $\mathcal{D}_{3000}^{\text{kSZ}}$  and  $\xi$  and  $\mathcal{D}_{3000}^{\text{tSZ}}$ , respectively. However, when removing the use of templates, these coefficients are reduced to 0.22 and  $-0.22$ , respectively, such that the level of correlation between the CIB and the tSZ is constrained to  $\xi = 0.07 \pm 0.03$ , ruling out positive correlations at the  $2\sigma$  level and confirming that there is an overdensity of dusty galaxies in galaxy clusters.



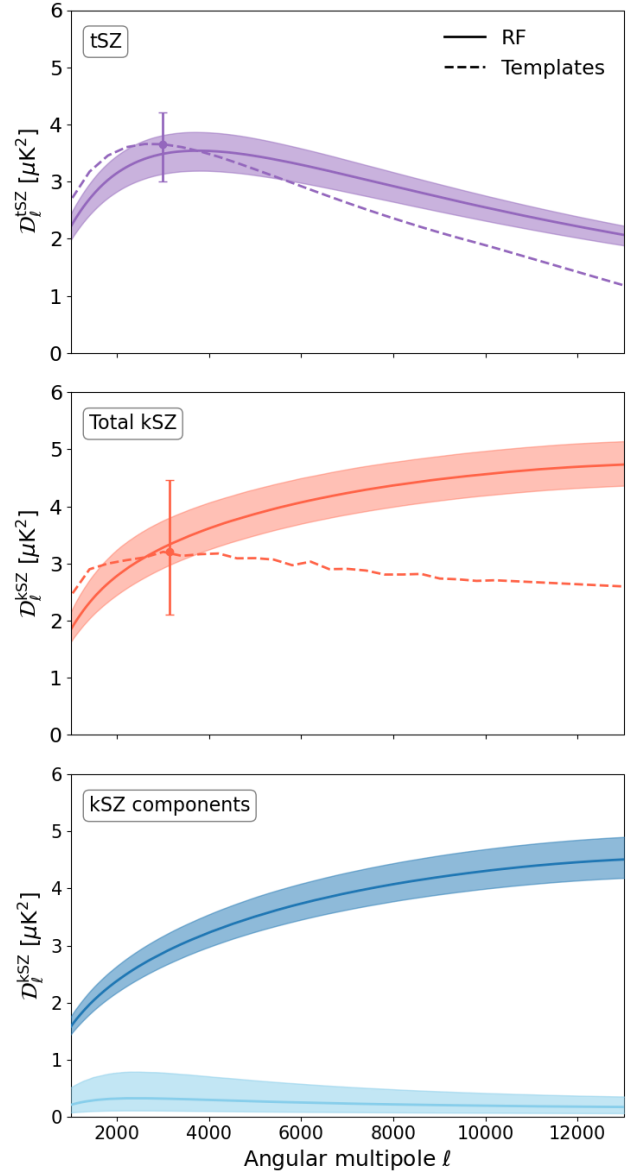


**Fig. 4.** Probability distributions of the amplitudes of the tSZ and kSZ power spectra at  $\ell = 3000$ , in  $\mu\text{K}^2$ , for the different models considered (see text). On the bottom panel, the kSZ is decomposed into its late-time and patchy components.

Regarding cosmological constraints, as shown in Table 3, we again recover the *Planck* priors on  $\Omega_b h^2$ ,  $\Omega_c h^2$ ,  $100\theta_{\text{MC}}$  and  $n_s$ . Overall, the values of the other sampled cosmological parameters are consistent with the large-scale results, only with larger error bars. Notably, the width of the 68% confidence interval on  $\log A$  increases from 0.014 to 0.027 when small-scale data are used in the analysis. In Fig. 7, we show the posterior distributions of  $A_s$ ,  $\tau$ ,  $z_{\text{re}}$  and  $z_{\text{end}}$  for our four cases. As before, we see that replacing the tSZ template by a cosmology-dependent spectrum does not impact the parameter constraints, because of the strong priors placed on  $\Omega_c h^2$  and so, effectively, on the shape of the spectrum. Introducing the RF-inferred kSZ spectrum, however, slightly shifts the most likely values of  $A_s$  and  $z_{\text{re}}$ , and, consequently, of  $\tau$ , pulling the constraints on the reionisation history towards later scenarios:  $z_{\text{re}}$  moves from  $8.4^{+0.8}_{-1.3}$  with both templates to  $z_{\text{re}} = 7.9^{+1.1}_{-1.3}$  with both cosmology-dependent spectra, whilst  $\tau$  moves from  $0.068^{+0.009}_{-0.014}$  to  $\tau = 0.062^{+0.012}_{-0.015}$  at the 68% confidence level.

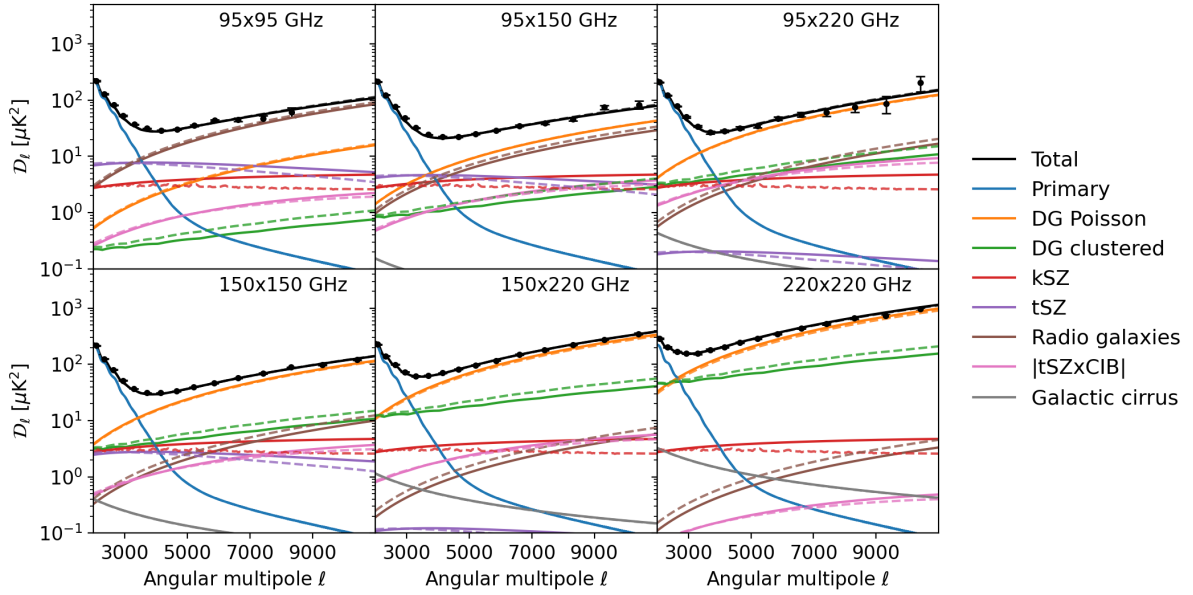
## 5. Discussion

The constraints on the history of reionisation derived from CMB analyses are traditionally based on the value of the Thomson optical depth which only offers a limited amount of information, since it is the integrated electron column density along the

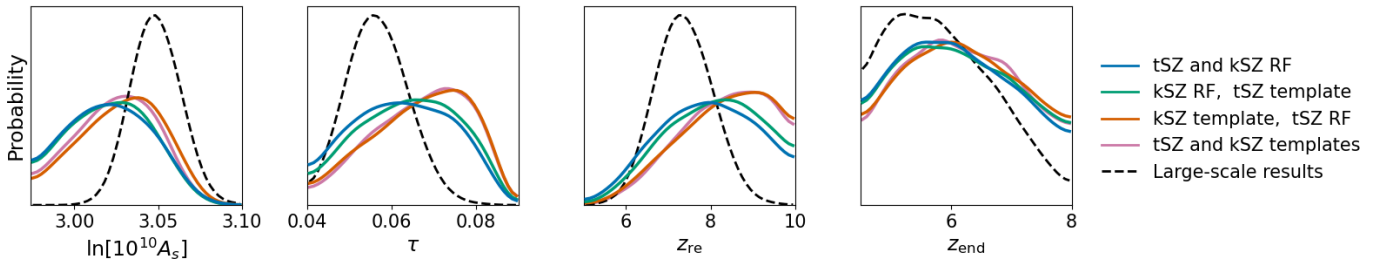


**Fig. 5.** Most likely SZ spectra at 150 GHz obtained when fitting cosmology and reionisation parameters to the SPT data. Two cases are compared: Either templates are used for both spectra (dashed lines) or they are predicted by our RF given the sampled parameters (solid lines). For the results obtained with RF predictions, shaded areas correspond to 68% confidence regions. For the results obtained with templates, only the amplitudes at  $\ell = 3000$  are constrained, and the 68% confidence intervals on their values are shown as error bars. The tSZ and total kSZ spectra are shown in the *upper and middle panels*, respectively. When possible, the kSZ signal is decomposed between patchy (dark blue) and late-time (light blue) components, shown in the *lower panel*.

line of sight, averaged over the whole sky. In particular, as seen in Sect. 4.1, constraining models with more than one parameter with current CMB data proves difficult. Additionally, the measurements of  $\tau$  from *Planck* data are still plagued with systematics and different analyses lead to different values (e.g., [Planck Collaboration XLVII 2016](#); [Pagano et al. 2020](#); [Natale et al. 2020](#)) which, whilst being mostly consistent, can change our picture of the high-redshift Universe. In order to improve the constraints on reionisation obtained from CMB data, one can combine CMB observations to other data sets, such as measurements of the IGM ionised fraction from quasar spectra, galaxy



**Fig. 6.** Most likely components of the small-scale CMB temperature power spectrum observed by the SPT (R21), when cosmology is free. Solid and dashed lines show the results when RF predictions and templates are used for the SZ spectra, respectively.



**Fig. 7.** Posterior distributions of cosmological and reionisation parameters sampled when cosmology is let free with [Planck Collaboration I \(2020\)](#) priors and RF-predicted or templates are used for the kSZ and tSZ power spectra. The constraints obtained with large-scale *Planck* data in Sect. 4.1 are shown as black dashed lines.

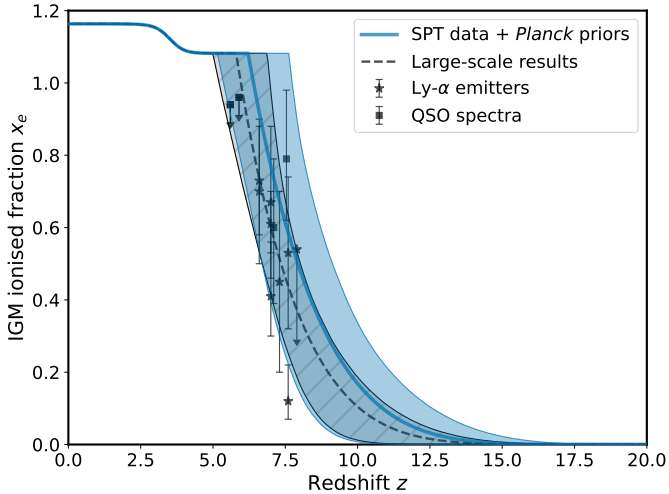
luminosity functions (e.g., [Gorce et al. 2018](#); [Qin et al. 2020](#)), or even measurements of the 21 cm signal (e.g., [Alvarez 2016](#); [Ahn & Shapiro 2021](#); [Paul et al. 2021](#); [Billings et al. 2021](#)). The option we chose in this work is to fully exploit the potential of CMB observations by including small-scale data in the analysis. In contrast to previous works ([Planck Collaboration XLVII 2016](#)), cosmological- and reionisation-dependent models of the tSZ and kSZ power spectra, which are foregrounds to the primary temperature fluctuations on scales  $\ell > 2000$ , are introduced, giving access to the cosmological information included in their shape and amplitude. Doing so, we provide new constraints on the nature of reionisation and on cosmological parameters.

### 5.1. Constraining the nature of reionisation

In Sect. 4.1, fitting the *Planck* data with different parameterisations of the redshift-evolution of the IGM ionised fraction (Eqs. (2) and (3)), we have found that the large-scale CMB power spectrum can only constrain the value of the optical depth  $\tau$ , but not the details of the reionisation history. However, the measured value of  $\tau$  can give a completely different picture of the high-redshift Universe, depending on the adopted model, and the latter should therefore be chosen carefully. A complementary analysis has been performed by [Qin](#)

et al. (2020), who directly confront semi-numerical simulations ([Mesinger et al. 2011](#); [Greig & Mesinger 2015](#); [Murray et al. 2020](#)) to [Planck Collaboration I \(2020\)](#) large-scale data, avoiding the use of any parameterisation of the reionisation history. They recover a constraint on  $\tau$  similar to parametric analyses, and argue that a prior on  $\tau$  can be used with the same effect as the full large-scale *Planck* likelihoods to constrain reionisation models. Another way of obtaining somewhat model-free constraints on reionisation from CMB data is the use of non-parametric reconstructions, for example via a principal component analysis ([Hu & Holder 2003](#); [Mortonson & Hu 2008](#); [Bégin et al. 2022](#)) or by interpolating between a varying number of redshift bins along the ionisation history ([Millea & Bouchet 2018](#)). These results motivate a more thorough analysis of small-scale data to complement large-scale results and better constrain reionisation.

Using cosmology-dependent SZ spectra to fit the small-scale SPT data, we find in Sect. 4.2 that their shape and amplitude can indeed be used to learn about the reionisation history. In Fig. 8, we compare the results for free cosmology runs on large-scale and small-scale data, and see that both data sets tell consistent stories, in agreement with current constraints on the ionisation fraction of the IGM from Lyman- $\alpha$  emitters ([Ota et al. 2008](#); [Schenker et al. 2013](#); [Konno et al. 2018](#); [Mason et al. 2018](#); [Hoag et al. 2019](#); [Jung et al. 2020](#)) and quasar spectra



**Fig. 8.** Constraints on reionisation history obtained with *Planck* (in grey, [Planck Collaboration I 2020](#)) and SPT high- $\ell$  data (in blue, [Reichardt et al. 2021](#)) for cosmology-dependent kSZ and tSZ spectra. Cosmology is free but, for the latter, Gaussian priors from Sect. 4.1 are imposed on four cosmological parameters. Inferences are compared to currently available data points on the ionisation fraction of the IGM from Lyman- $\alpha$  emitters and QSO spectra (see text for references).

([McGreer et al. 2015](#); [Greig et al. 2017, 2019](#)), or a combination of both ([Gorce et al. 2018](#)). With large-scale data only, we report a constraint on the midpoint of reionisation of  $7.4 \pm 0.7$ . Using the cosmology-dependent kSZ and tSZ spectra to analyse the SPT data ([Reichardt et al. 2021](#)), we obtain  $z_{\text{re}} = 7.9^{+1.1}_{-1.3}$  for the midpoint and  $z_{\text{end}} = 6.2^{+1.6}_{-0.8}$ , corresponding to a 95% upper limit of  $z_{\text{end}} < 8.5$ , for the endpoint (see Fig. 7). The small-scale data requires larger values of the reionisation midpoint than large-scale results, with the first light sources starting to ionise the IGM around  $z = 15$ <sup>8</sup>. This fact was already observed in [Choudhury et al. \(2021\)](#), where the authors fit a semi-numerical model of reionisation to *Planck* CMB and SPT data and could mean a tension between the small-scale CMB data and recent works inferring a late end for reionisation from astrophysical data sets ([Becker et al. 2015](#); [Bosman et al. 2018](#)). More precisely, let the duration of reionisation  $\Delta_{z,50} = z_{0.25} - z_{0.75}$ , where  $z_i$  is the redshift at which a fraction  $i$  of the IGM is ionised. We find an upper limit  $\Delta_{z,50} < 4.8$  (4.3) at 99% (95%) confidence, consistent with the constraints obtained in previous works from the SPT measurement of the kSZ amplitude ([Reichardt et al. 2021](#)). These include  $\Delta_{z,50} < 2.9$  (99% C.L.) presented in [Choudhury et al. \(2021\)](#) and  $\Delta_{z,50} < 3.5$  given in [Chen et al. \(2022\)](#), although we expect the latter limit to be significantly loosened when the parameters of the model used by the authors are freed. Our upper limit is also less stringent than the initial results of [Planck Collaboration XLVII \(2016\)](#), where the authors post-process the posterior distribution of  $\mathcal{D}_{3000}^{\text{kSZ}}$  obtained with templates confronted to a combination of *Planck*, ACT and SPT data ([Das et al. 2014](#); [George et al. 2015](#); [Planck Collaboration I 2016](#)). We add that, as mentioned in Appendix A, kSZ measurements alone seem insufficient to obtain precise constraints on  $z_{\text{re}}$ , and a coherent analysis of small- and large-scale data is required.

The new analytic derivation of the kSZ angular spectrum introduced in [Gorce et al. \(2020\)](#) and confronted to data for the

first time in this work does not only allow a consistent modelling of the reionisation history in large and small CMB data analysis, but also gives access to information about the patchiness of reionisation and the nature of its sources. Indeed, the  $\kappa$  parameter is directly related to the typical size of ionised bubbles, whilst  $\alpha_0$  describes the variance of the electron density field during the EoR, and the distribution of light sources throughout the IGM. However, current data does not allow to get precise constraints on these two shape parameters. We can report an upper limit  $\alpha_0 < 10^{4.1} \text{ Mpc}^3$  at 95% confidence, which is in good agreement with values obtained for both radiative hydrodynamics and semi-numerical simulations of reionisation ([Mesinger et al. 2011](#); [Aubert et al. 2015](#); [Seiler et al. 2019](#)) in [Gorce et al. \(2020\)](#). We expect the results presented here to greatly improve with data from the next generation of CMB observatories. Additionally, these parameters being related to the morphology of reionisation, and so to the physical properties of early galaxies leading the process, future observations, for example of the spatial fluctuations of the 21 cm signal by the Square Kilometre Array ([Mellema et al. 2015](#)), will give access to independent constraints on these parameters, and therefore help the characterisation of the high-multipole CMB power spectrum.

## 5.2. Cosmological constraints

In Fig. 7, we compare the posterior distributions of  $\log A$ ,  $\tau$ ,  $z_{\text{re}}$  and  $z_{\text{end}}$  obtained in Sect. 4.1 with *Planck* data only, and in Sect. 4.2 with the SPT data, *Planck* priors and RF-inferred SZ spectra. We see that, because of the presence of many different components to the observed temperature power spectrum on small scales, the SPT data are less constraining than *Planck* CMB. In particular, the  $1\sigma$  intervals on  $\log A$  and  $\tau$  close to double, from 0.007 reported in Sect. 4.1 to 0.012 for the latter. Interestingly, the  $\tau$ - $A_s$  degeneracy is not broken by the new reionisation model and the addition of the SPT data. However, we expect that future experiments will be able to break this degeneracy ([Calabrese et al. 2014](#); [Galli et al. 2014](#)) and such forecasts will be the focus of future work. The degeneracy between the two parameters, although corresponding to the same correlation coefficient, is shifted between the large- and small-scale results. That is, for a given  $\log A$ , large-scale data favours larger  $\tau$  values than small-scale data, and reciprocally. This is due to the small-scale data requiring a kSZ amplitude corresponding to early reionisation scenarios, which in turn translates into a larger optical depth. Hence, the best-fit value for  $\log A$  is found to be smaller than, although still fully consistent with, the large-scale results.

Another potential issue of this analysis is the strong correlation observed between the reionisation midpoint  $z_{\text{re}}$  and the Thomson optical depth  $\tau$ , which is directly derived from the reionisation history. Indeed, we report a correlation coefficient between the two parameters of 0.98. This degeneracy could be partly broken with future data to allow for a cleaner measurement of both reionisation history parameters,  $z_{\text{re}}$  and  $z_{\text{end}}$ . For example, measurements of the spatial fluctuations of the CMB optical depth could be used ([Holder et al. 2007](#); [Dvorkin & Smith 2009](#); [Gluscevic et al. 2013](#); [Namikawa et al. 2021](#)). Another indirect option would be to independently constrain  $z_{\text{re}}$  and  $z_{\text{end}}$  with data from other cosmological sources, such as quasar spectra or the power spectrum of fluctuations in the 21 cm signal coming from the EoR ([Liu et al. 2016](#)) or with next generation CMB data. Indeed, [Alvarez et al. \(2021\)](#) suggest that varying the reionisation history has a different impact on the

<sup>8</sup> The parameterisation used allows for reionisation to start as early as  $z = 20$ , but not before.

Gaussian and the non-Gaussian part of the signal<sup>9</sup>, and, therefore, four-point estimators of the kSZ could be used to break the degeneracy.

In order to break degeneracies between parameters, detect systematics, and allow a fruitful analysis of available data sets, a consistent analysis of the different observables of reionisation seems to be a promising avenue. Indeed, several works have already demonstrated the potential offered by a combined analysis of astrophysical and CMB data (e.g., Millea & Bouchet 2018; Qin et al. 2020; Chatterjee et al. 2021; Baxter et al. 2021; Sato-Polito et al. 2021; La Plante et al. 2022), 21 cm and CMB data (e.g., Tashiro et al. 2011; Alvarez 2016; Ma et al. 2018; Hotinli & Johnson 2022; La Plante et al. 2020; Billings et al. 2021; Choudhury et al. 2021; Ahn & Shapiro 2021; Bégin et al. 2022), or different types of CMB data (e.g., Planck Collaboration XLVII 2016; Alvarez et al. 2021; Namikawa et al. 2021; Paul et al. 2021; Roy et al. 2021).

## 6. Conclusions

Using machine learning, we have included a cosmology-dependent derivation of the tSZ (following, among others, Komatsu & Seljak 2002; Taburet et al. 2009; Planck Collaboration XVI 2014; Salvati et al. 2018) and kSZ angular power spectra, introduced in Gorce et al. (2020), to the analysis of the SPT measurements of the small-scale power spectrum of CMB temperature fluctuations. Doing so, we have proven that the cosmological information included in the shape and amplitude of these foregrounds, usually considered as nuisance parameters, can help tighten constraints on cosmology and reionisation. In Douspis et al. (2022), we have focused our efforts on the tSZ effect. In this work, we have additionally included a proper modelling of reionisation and of the kSZ effect to the analysis.

Several works have demonstrated that the history of reionisation is not an instantaneous, redshift-symmetric process, motivating the inclusion of a physical parameterisation of reionisation (Douspis et al. 2015) in the analysis of the latest *Planck* CMB data (Planck Collaboration I 2020). In the framework introduced in this work, it is not the Thomson optical depth  $\tau$  that is sampled, but the reionisation history itself, through two meaningful parameters: The reionisation midpoint  $z_{\text{re}}$  and endpoint  $z_{\text{end}}$ , corresponding to the redshift when the hydrogen in the IGM is 50% and 100% ionised, respectively. Analysing the *Planck* data, in Sect. 4.1, we find a value of the Thomson optical depth  $\tau = 0.056 \pm 0.07$  fully consistent with the one obtained using the fiducial, symmetric model ( $\tau = 0.054 \pm 0.007$ ). However, the reionisation history corresponding to this value gives a completely different picture of the high-redshift Universe, where the first light sources light up and start reionising the IGM as soon as  $z = 15$  (compared to  $z = 10$  for the instantaneous model, see Fig. 2). Indeed, despite the  $A_s$ - $\tau$  degeneracy translating into a  $A_s$ - $z_{\text{re}}$  degeneracy, the *Planck* data alone gives tight constraints on the reionisation history, with a midpoint  $z_{\text{re}} = 7.4 \pm 0.7$  and an upper limit on the end of the process  $z_{\text{end}} < 7.5$  (95% C.L.). These results corroborate previous works (e.g. Planck Collaboration XLVII 2016; Qin et al. 2020), who found that the large-scale CMB data was really only sensitive to the integrated reionisation history, through  $\tau$ , and not to the details of the process. They motivate a new analysis of the small-scale CMB temperature power spectrum, designed to

exploit the cosmological information included not only in the primary signal, but also in the CMB foregrounds, as done in Sect. 4.2.

In the analysis of the SPT small-scale data (Reichardt et al. 2021), we replace the template of the kSZ spectrum used in previous analyses by a cosmology-dependent estimate (Fig. 1). Because a full derivation at each iteration of the sampler is too time-consuming to be considered, we train random forests to predict the patchy and late-time kSZ angular power spectra, when given the following parameters:  $\Omega_b h^2$ ,  $\Omega_c h^2$ ,  $100\theta_{\text{MC}}$ ,  $n_s$ ,  $A_s$ ,  $z_{\text{re}}$ ,  $z_{\text{end}}$ ,  $\alpha_0$  and  $\kappa$  – the latter two defining the shape of the patchy kSZ spectrum (Gorce et al. 2020)<sup>10</sup>. We incorporate the large-scale results to the small-scale analysis by imposing Gaussian priors corresponding to the results of Sect. 4.1 on  $\Omega_b h^2$ ,  $\Omega_c h^2$ ,  $100\theta_{\text{MC}}$  and  $n_s$ . We find that removing the tSZ and kSZ templates from the analysis, with or without letting the cosmology free, breaks the degeneracy between the amplitudes at  $\ell = 3000$  of the two power spectra. We report a  $5\sigma$  measurement of the amplitude of the kSZ angular power spectrum, marginalised over cosmology (see Fig. 4):

$$\mathcal{D}_{3000}^{\text{kSZ}} = 3.4_{-0.3}^{+0.5} \mu\text{K}^2 \text{ (68\% C.L.)}.$$

The previous measurement, obtained with the same data but fixed cosmology, was given at  $3\sigma$  (Reichardt et al. 2021). We confirm their result of a negative tSZxCIB correlation, with a zero correlation being disfavoured above the  $2\sigma$  level. These values result in a well-constrained reionisation history, although in slight tension with *Planck*: the SPT data indeed seem to favour earlier reionisation scenarios, with (see Fig. 7)

$$z_{\text{re}} = 7.9_{-1.3}^{+1.1} \text{ (68\% C.L.)},$$

leading to an optical depth

$$\tau = 0.062_{-0.015}^{+0.012} \text{ (68\% C.L.)}.$$

These scenarios are in agreement with current constraints on the IGM ionisation level obtained from Lyman- $\alpha$  emitters and quasar spectra (Fig. 8). We also report the first direct upper limit on the amplitude of the kSZ power stemming from reionisation

$$\mathcal{D}_{3000}^{\text{pkSZ}} < 1.6 \mu\text{K}^2 \text{ (95\% C.L.)},$$

corresponding to about half of the previous upper limit (Reichardt et al. 2021) and show in Fig. 5, for the first time, the shape of the SZ spectra preferred by the small-scale CMB data. In particular, these new results seem to indicate a stronger small-scale power of both the tSZ and kSZ effects than previously allowed by templates.

If these results are promising, they can still be improved. In future work, we will focus our efforts on the modelling of other high-multipole CMB foregrounds, such as the cosmic infrared background (Maniyar et al. 2019, 2021), and perform a simultaneous analysis of small- and large-scale CMB data in a unified framework to achieve better constraints on the reionisation history. We will also explore the potential of the future generation of CMB observatories to measure the shape of the kSZ angular power spectrum<sup>11</sup>, and, doing so, to constrain the morphology of reionisation<sup>12</sup>.

<sup>10</sup> The trained random forests are available at <https://szdb.osups.universite-paris-saclay.fr>.

<sup>11</sup> See <http://idoc.ias.u-psud.fr>.

<sup>12</sup> See <http://www.dimacav-plus.fr>.

<sup>9</sup> For example, the kSZ four-point correlations will be enhanced by shorter reionisation histories.



**Table 3.** Maximum likelihood parameters obtained with SPT data (Reichardt et al. 2021) when RF predictions or templates are used for the SZ effects and templates for the other foregrounds.

Parameter	Priors	kSZ and tSZ templates	Cosmology-dependent kSZ and tSZ
$\Omega_b h^2$	$0.0224 \pm 0.0001$	id.	id.
$\Omega_c h^2$	$0.120 \pm 0.001$	id.	id.
$100\theta_{MC}$	$1.0409 \pm 0.0003$	id.	id.
$n_s$	$0.965 \pm 0.004$	id.	id.
$\ln(10^{10} A_s)$	$2.97 - 3.12$	$3.026^{+0.022}_{-0.028}$	$3.021^{+0.026}_{-0.028}$
$1 - b$	$0.5 - 1.0$	–	$0.57 \pm 0.03$
$z_{re}$	$5 - 10$	$8.4^{+0.8}_{-1.3}$	$7.9^{+1.1}_{-1.3}$
$dz$	$0 - 4$	$1.8^{+1.3}_{-0.8}$	$1.4^{+1.2}_{-0.6}$
$\log[\alpha_0/\text{Mpc}^3]$	$2.5 - 4.5$	–	$<4.1$
$\kappa [\text{Mpc}^{-1}]$	$0.04 - 0.20$	–	N.A.
$\tau$	–	$0.068^{+0.009}_{-0.014}$	$0.062^{+0.012}_{-0.015}$
$z_{end}$	$>4.5$	$<8.6$	$<8.5$
$\mathcal{D}^{tSZ}$	–	$3.7 \pm 0.5$	$3.5 \pm 0.4$
$\mathcal{D}^{kSZ}$	–	$3.1 \pm 1.0$	$3.4^{+0.5}_{-0.3}$
Patchy	–	–	$<1.6$
Late-time	–	–	$2.9^{+0.3}_{-0.2}$

**Notes.** Priors are either Gaussian or flat. In particular, Gaussian priors corresponding to  $1\sigma$  best-fit results from Planck Collaboration I (2020) are imposed on cosmological parameters. The lower part of the table corresponds to derived parameters. Errors and upper limits correspond to 68% and 95% confidence intervals, respectively. Spectrum amplitudes are given in  $\mu\text{K}^2$  and at  $\ell = 3000$ . The last column presents the main results of this paper.

**Acknowledgements.** The authors thank Dominique Aubert, Ian Hothi, Stéphane Ilić and Adrian Liu for useful comments regarding a draft version of this paper. They additionally thank the referee for their useful comments which certainly helped improve this work. AG’s work was supported by the McGill Astrophysics Fellowship funded by the Trottier Chair in Astrophysics, as well as the Canadian Institute for Advanced Research (CIFAR) Azrieli Global Scholars program and the Canada 150 Programme. This work was additionally supported by the Programme National Cosmologie et Galaxies (PNCG) of CNRS/INSU with INP and IN2P3, co-funded by CEA and CNES. Most of the analysis was performed at the computation facility of IDOC, partly provided by DIM ACAV. The authors also acknowledge the use of the supercomputer Cedar at Simon Fraser University, managed by Compute Canada to compute the training sets used in this work. The operation of this supercomputer is funded by the Canada Foundation for Innovation (CFI). This research made use of `matplotlib`, a Python library for publication quality graphics (Hunter 2007), of `scipy`, a Python-based ecosystem of open-source software for mathematics, science, and engineering (Jones et al. 2001) – including `numpy` (Oliphant 2006), of the `scikit-learn` (Pedregosa et al. 2011) library, and of the corner (Foreman-Mackey 2016), `getdist` (Lewis 2019) and `arviz` (Kumar et al. 2019) packages to analyse posterior distributions.

## References

- Abazajian, K. N., Adshead, P., Ahmed, Z., et al. 2016, *ApJ*, submitted [arXiv:1610.02743]
- Addison, G. E., Dunkley, J., & Spergel, D. N. 2012, *MNRAS*, 427, 1741
- Ade, P., Aguirre, J., Ahmed, Z., et al. 2019, *J. Cosmol. Astropart. Phys.*, 2019, 056
- Ahn, K., & Shapiro, P. R. 2021, *ApJ*, 914, 44
- Alvarez, M. A. 2016, *ApJ*, 824, 118
- Alvarez, M. A., Ferraro, S., Hill, J. C., Hložek, R., & Ikape, M. 2021, *Phys. Rev. D*, 103, 063518
- An, L., Brooks, S., & Gelman, A. 1998, *J. Comput. Graph. Stat.*, 7, 434
- Aubert, D., Deparis, N., & Ocvirk, P. 2015, *MNRAS*, 454, 1012
- Battaglia, N., Natarajan, A., Trac, H., Cen, R., & Loeb, A. 2013, *ApJ*, 776, 83
- Baxter, E. J., Weinberger, L., Haehnelt, M., et al. 2021, *MNRAS*, 501, 6215
- Becker, G. D., Bolton, J. S., & Lidz, A. 2015, *PASA*, 32, e045
- Bégin, J.-M., Liu, A., & Gorce, A. 2022, *Phys. Rev. D*, 105
- Bennett, C. L., Banday, A. J., Gorski, K. M., et al. 1996, *ApJ*, 464, L1
- Billings, T. S., La Plante, P., & Aguirre, J. E. 2021, *PASP*, submitted [arXiv:2103.14563]
- Bosman, S. E. I., Fan, X., Jiang, L., et al. 2018, *MNRAS*, 479, 1055
- Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015, *ApJ*, 811, 140
- Calabrese, E., Hložek, R., Battaglia, N., et al. 2014, *J. Cosmol. Astropart. Phys.*, 2014, 010
- Casey, C. M., Narayanan, D., & Cooray, A. 2014, *Phys. Rep.*, 541, 45
- Chatterjee, A., Choudhury, T. R., & Mitra, S. 2021, *MNRAS*, 507, 2405
- Chen, N., Trac, H., Mukherjee, S., & Cen, R. 2022, *ApJ*, submitted [arXiv:2203.04337]
- Choudhury, T. R., Mukherjee, S., & Paul, S. 2021, *MNRAS*, 501, L7
- Das, S., Louis, T., Nolta, M. R., et al. 2014, *J. Cosmol. Astropart. Phys.*, 2014, 014
- Douspis, M., Aghanim, N., & Langer, M. 2006, *A&A*, 456, 819
- Douspis, M., Aghanim, N., Ilić, S., & Langer, M. 2015, *A&A*, 580, L4
- Douspis, M., Salvati, L., Gorce, A., & Aghanim, N. 2022, *A&A*, 659, A99
- Dunkley, J., Calabrese, E., Sievers, J., et al. 2013, *J. Cosmol. Astropart. Phys.*, 2013, 025
- Dvorkin, C., & Smith, K. M. 2009, *Phys. Rev. D*, 79, 043003
- Foreman-Mackey, D. 2016, *J. Open Source Softw.*, 1, 24
- Galli, S., Benabed, K., Bouchet, F., et al. 2014, *Phys. Rev. D*, 90, 063504
- Gelman, A., & Rubin, D. B. 1992, *Stat. Sci.*, 7, 457
- George, E. M., Reichardt, C. L., Aird, K. A., et al. 2015, *ApJ*, 799, 177
- Gluscevic, V., Kamionkowski, M., & Hanson, D. 2013, *Phys. Rev. D*, 87, 047303
- Gorce, A., Douspis, M., Aghanim, N., & Langer, M. 2018, *A&A*, 616, A113
- Gorce, A., Ilić, S., Douspis, M., Aubert, D., & Langer, M. 2020, *A&A*, 640, A90
- Greig, B., & Mesinger, A. 2015, *MNRAS*, 449, 4246
- Greig, B., Mesinger, A., Haiman, Z., & Simcoe, R. A. 2017, *MNRAS*, 466, 4239
- Greig, B., Mesinger, A., & Bañados, E. 2019, *MNRAS*, 484, 5094
- Hoag, A., Bradač, M., Huang, K., et al. 2019, *ApJ*, 878, 12
- Holder, G. P., Iliev, I. T., & Mellema, G. 2007, *ApJ*, 663, L1
- Hotinli, S. C., & Johnson, M. C. 2022, *Phys. Rev. D*, 105, 063522
- Howlett, C., Lewis, A., Hall, A., & Challinor, A. 2012, *J. Cosmol. Astropart. Phys.*, 1204, 027
- Hu, W., & Holder, G. P. 2003, *Phys. Rev. D*, 68, 023001
- Hunter, J. D. 2007, *Comput. Sci. Eng.*, 9, 90
- Ichiki, K. 2014, *Progr. Theoret. Exp. Phys.*, 2014, 06B109
- Iliev, I. T., Pen, U.-L., Bond, J. R., Mellema, G., & Shapiro, P. R. 2007, *ApJ*, 660, 933
- Jones, E., Oliphant, T., Peterson, P., et al. 2001, *SciPy: Open Source Scientific Tools for Python*
- Jung, I., Finkelstein, S. L., Dickinson, M., et al. 2020, *ApJ*, 904, 144
- Komatsu, E., & Seljak, U. 2002, *MNRAS*, 336, 1256
- Konno, A., Ouchi, M., Shibuya, T., et al. 2018, *PASJ*, 70, S16

- Kosowsky, A. 2003, *New Astron. Rev.*, **47**, 939
- Kuhlen, M., & Faucher-Giguère, C.-A. 2012, *MNRAS*, **423**, 862
- Kumar, R., Carroll, C., Hartikainen, A., & Martin, O. 2019, *J. Open Source Softw.*, **4**, 1143
- La Plante, P., Lidz, A., Aguirre, J., & Kohn, S. 2020, *ApJ*, **899**, 40
- La Plante, P., Sipple, J., & Lidz, A. 2022, *ApJ*, **928**, 162
- Lagache, G., Puget, J.-L., & Dole, H. 2005, *ARA&A*, **43**, 727
- Lewis, A. 2013, *Phys. Rev. D*, **87**, 103529
- Lewis, A. 2019, ArXiv e-prints [arXiv:1910.13970]
- Lewis, A., & Bridle, S. 2002, *Phys. Rev. D*, **66**, 103511
- Lewis, A., Challinor, A., & Lasenby, A. 2000, *ApJ*, **538**, 473
- Liu, A., Pritchard, J. R., Allison, R., et al. 2016, *Phys. Rev. D*, **93**, 043013
- Ma, Q., Helgason, K., Komatsu, E., Ciardi, B., & Ferrara, A. 2018, *MNRAS*, **476**, 4025
- Maniyar, A., Lagache, G., Béthermin, M., & Ilić, S. 2019, *A&A*, **621**, A32
- Maniyar, A., Béthermin, M., & Lagache, G. 2021, *A&A*, **645**, A40
- Mason, C. A., Treu, T., Dijkstra, M., et al. 2018, *ApJ*, **856**, 2
- McGreer, I. D., Mesinger, A., & D'Odorico, V. 2015, *MNRAS*, **447**, 499
- McQuinn, M., Furlanetto, S. R., Hernquist, L., Zahn, O., & Zaldarriaga, M. 2005, *ApJ*, **630**, 643
- Mellema, G., Koopmans, L., Shukla, H., et al. 2015, in *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, 10
- Mesinger, A., Furlanetto, S., & Cen, R. 2011, *MNRAS*, **411**, 955
- Mesinger, A., McQuinn, M., & Spergel, D. N. 2012, *MNRAS*, **422**, 1403
- Millea, M., & Bouchet, F. 2018, *A&A*, **617**, A96
- Mortonson, M. J., & Hu, W. 2008, *ApJ*, **672**, 737
- Murray, S. G., Greig, B., Mesinger, A., et al. 2020, *J. Open Source Softw.*, **5**, 2582
- Namikawa, T., Roy, A., Sherwin, B. D., Battaglia, N., & Spergel, D. N. 2021, *Phys. Rev. D*, **104**, 063514
- Natale, U., Pagano, L., Lattanzi, M., et al. 2020, *A&A*, **644**, A32
- Oliphant, T. 2006, *NumPy: A guide to NumPy* (USA: Trelgol Publishing)
- Ota, K., Iye, M., Kashikawa, N., et al. 2008, *ApJ*, **677**, 12
- Pagano, L., Delouis, J. M., Mottet, S., Puget, J. L., & Vibert, L. 2020, *A&A*, **635**, A99
- Park, H., Shapiro, P. R., Komatsu, E., et al. 2013, *ApJ*, **769**, 93
- Park, J., Mesinger, A., Greig, B., & Gillet, N. 2019, *MNRAS*, **484**, 933
- Paul, S., Mukherjee, S., & Choudhury, T. R. 2021, *MNRAS*, **500**, 232
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *J. Mach. Learn. Res.*, **12**, 2825
- Planck Collaboration XVI. 2014, *A&A*, **571**, A21
- Planck Collaboration I. 2016, *A&A*, **594**, A1
- Planck Collaboration XLVII. 2016, *A&A*, **596**, A108
- Planck Collaboration I. 2020, *A&A*, **641**, A1
- Planck Collaboration VI. 2020, *A&A*, **641**, A6
- Qin, Y., Poulin, V., Mesinger, A., et al. 2020, *MNRAS*, **499**, 550
- Reichardt, C. L., Shaw, L., Zahn, O., et al. 2012, *ApJ*, **755**, 70
- Reichardt, C. L., Patil, S., Ade, P. A. R., et al. 2021, *ApJ*, **908**, 199
- Robertson, B. E., Ellis, R. S., Furlanetto, S. R., & Dunlop, J. S. 2015, *ApJ*, **802**, L19
- Roy, A., Kulkarni, G., Meerburg, P. D., et al. 2021, *J. Cosmol. Astropart. Phys.*, **2021**, 003
- Ruhl, J., Ade, P. A. R., Carlstrom, J. E., et al. 2004, in *Z-Spec: a Broadband Millimeter-wave Grating Spectrometer: Design, Construction, and First Cryogenic Measurements*, eds. C. M. Bradford, P. A. R. Ade, J. E. Aguirre, et al., *SPIE Conf. Ser.*, **5498**, 11
- Salvati, L., Douspis, M., & Aghanim, N. 2018, *A&A*, **614**, A13
- Salvati, L., Douspis, M., Ritz, A., Aghanim, N., & Babul, A. 2019, *A&A*, **626**, A27
- Sato-Polito, G., Bernal, J. L., Boddy, K. K., & Kamionkowski, M. 2021, *Phys. Rev. D*, **103**, 083519
- Schenker, M. A., Robertson, B. E., Ellis, R. S., et al. 2013, *ApJ*, **768**, 196
- Seiler, J., Hutter, A., Sinha, M., & Croton, D. 2019, *MNRAS*, **1578**
- Shang, C., Haiman, Z., Knox, L., & Oh, S. P. 2012, *MNRAS*, **421**, 2832
- Shaw, L. D., Nagai, D., Bhattacharya, S., & Lau, E. T. 2010, *ApJ*, **725**, 1452
- Shaw, L. D., Rudd, D. H., & Nagai, D. 2012, *ApJ*, **756**, 15
- Sievers, J. L., Hlozek, R. A., Nolta, M. R., et al. 2013, *J. Cosmol. Astropart. Phys.*, **2013**, 060
- Sobrin, J. A., Anderson, A. J., Bender, A. N., et al. 2022, *ApJS*, **258**, 42
- Sunyaev, R. A., & Zeldovich, I. B. 1980, *ARA&A*, **18**, 537
- Taburet, N., Aghanim, N., Douspis, M., & Langer, M. 2009, *MNRAS*, **392**, 1153
- Tanimura, H., Douspis, M., Aghanim, N., & Salvati, L. 2022, *MNRAS*, **509**, 300
- Tashiro, H., Aghanim, N., Langer, M., et al. 2011, *MNRAS*, **414**, 3424
- Trac, H., Chen, N., Holst, I., Alvarez, M. A., & Cen, R. 2022, *ApJ*, **927**, 186
- Viero, M. P., Wang, L., Zemcov, M., et al. 2013, *ApJ*, **772**, 77
- Zahn, O., Lidz, A., McQuinn, M., et al. 2007, *ApJ*, **654**, 12
- Zahn, O., Mesinger, A., McQuinn, M., et al. 2011, *MNRAS*, **414**, 727
- Zahn, O., Reichardt, C. L., Shaw, L., et al. 2012, *ApJ*, **756**, 65
- Zeldovich, Y. B., & Sunyaev, R. A. 1969, *Astrophys. Space Sci.*, **4**, 301

## Appendix A: Random forests predictions

The derivation of the kSZ angular power spectrum introduced in Gorce et al. (2020) is too time-consuming to be included as such in a sampling algorithm. For this reason, and as already successfully implemented in Paper I to estimate the tSZ power, we train random forests to approximate the kSZ angular power spectrum, given a set of cosmological and reionisation parameters. We refer the reader to Paper I for more details about our choice of machine learning technique and parameter optimisation.

We use for training a set of 50 000 kSZ power spectra computed on 25  $\ell$ -bins ( $100 \leq \ell \leq 10\,500$ )<sup>13</sup> for parameters  $\ln(10^{10} A_s)$ ,  $\Omega_b h^2$ ,  $\Omega_c h^2$ ,  $100 \theta_{MC}$ ,  $n_s$ ,  $z_{re}$ ,  $dz$ ,  $\log \alpha_0$  and  $\kappa$ . The parameter ranges covered by the training set include the sampling ranges of Sec. 3.2. The samples are divided into fifths such that 60%, 20% and 20% of the models are used for the learning, validation, and testing phase, respectively. Because of their differences in shape and amplitude, as well as their sensitivity to different cosmological parameters, we train two separate sets of random forests to predict the late-time and patchy kSZ signals.

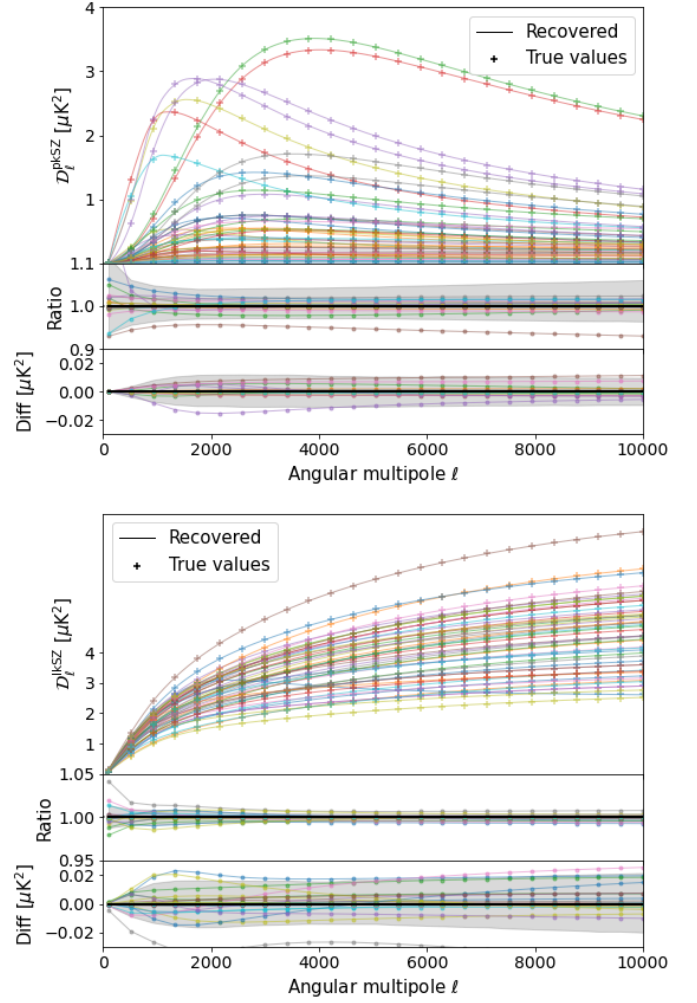
In order to facilitate the training, we scale each spectrum in the training set by  $\mathcal{D}_{3000}$  values obtained with the method described below. Doing so, the amplitude differences between the spectra in the training set will be smoothed out, and the random forests will focus their training on the dependency between parameters and the shape of the spectrum, rather than its amplitude. We use an MCMC sampling algorithm to obtain the parameters  $\{\alpha_i\}$  that minimise the distance between the true  $\mathcal{D}_{3000}^{true}$  and a model given by  $\mathcal{D}_{3000}^{model} \equiv A_0 \prod_i [\theta_i / \theta_{ref,i}]^{\alpha_i}$ , for each of the cosmological and reionisation parameter sets  $\{\theta_i\}$  in the training set. We fix  $A_0$  to the amplitude of the spectrum at  $\ell = 3000$  obtained with the full computation for  $\{\theta_{ref,i}\}$ . We obtain the following for the patchy kSZ amplitude:

$$\begin{aligned} \mathcal{D}_{3000}^{pkSZ} &= 1.03 \mu K^2 \times \left(\frac{z_{re}}{7.36}\right)^{0.70} \left(\frac{dz}{1.4}\right)^{0.87} \left(\frac{\log \alpha_0}{3.7}\right)^{6.60} \left(\frac{\kappa}{0.10}\right)^{1.99} \\ &\times \left(\frac{\Omega_b h^2}{0.0224}\right)^{1.47} \left(\frac{\Omega_c h^2}{0.120}\right)^{0.20} \left(\frac{n_s}{0.9653}\right)^{-0.28} \\ &\times \left(\frac{100\theta_{MC}}{1.041}\right)^{0.012} \left(\frac{\log A}{3.048}\right)^{2.18}, \end{aligned} \quad (A.1)$$

and the following for the late-time kSZ amplitude:

$$\begin{aligned} \mathcal{D}_{3000}^{lkSZ} &= 3.11 \mu K^2 \times \left(\frac{z_{re}}{7.37}\right)^{0.24} \left(\frac{dz}{1.4}\right)^{-0.08} \left(\frac{\log \alpha_0}{3.7}\right)^{0.07} \left(\frac{\kappa}{0.10}\right)^{0.04} \\ &\times \left(\frac{\Omega_b h^2}{0.0224}\right)^{0.78} \left(\frac{\Omega_c h^2}{0.120}\right)^{2.93} \left(\frac{n_s}{0.9653}\right)^{2.12} \\ &\times \left(\frac{100\theta_{MC}}{1.041}\right)^{-0.006} \left(\frac{\log A}{3.048}\right)^{6.94}. \end{aligned} \quad (A.2)$$

We note that the observed scalings, although model-dependent, exhibit the expected dependencies. For example, the amplitude of the late-time kSZ is inversely proportional to the duration of reionisation,  $dz$ , but proportional to the reionisation redshift  $z_{re}$ . Indeed, for a given  $z_{re}$ , a longer reionisation will end later, and, therefore, the late-time signal will be integrated over a short redshift range. Conversely, for a given reionisation duration, a larger midpoint means that reionisation will end earlier and the

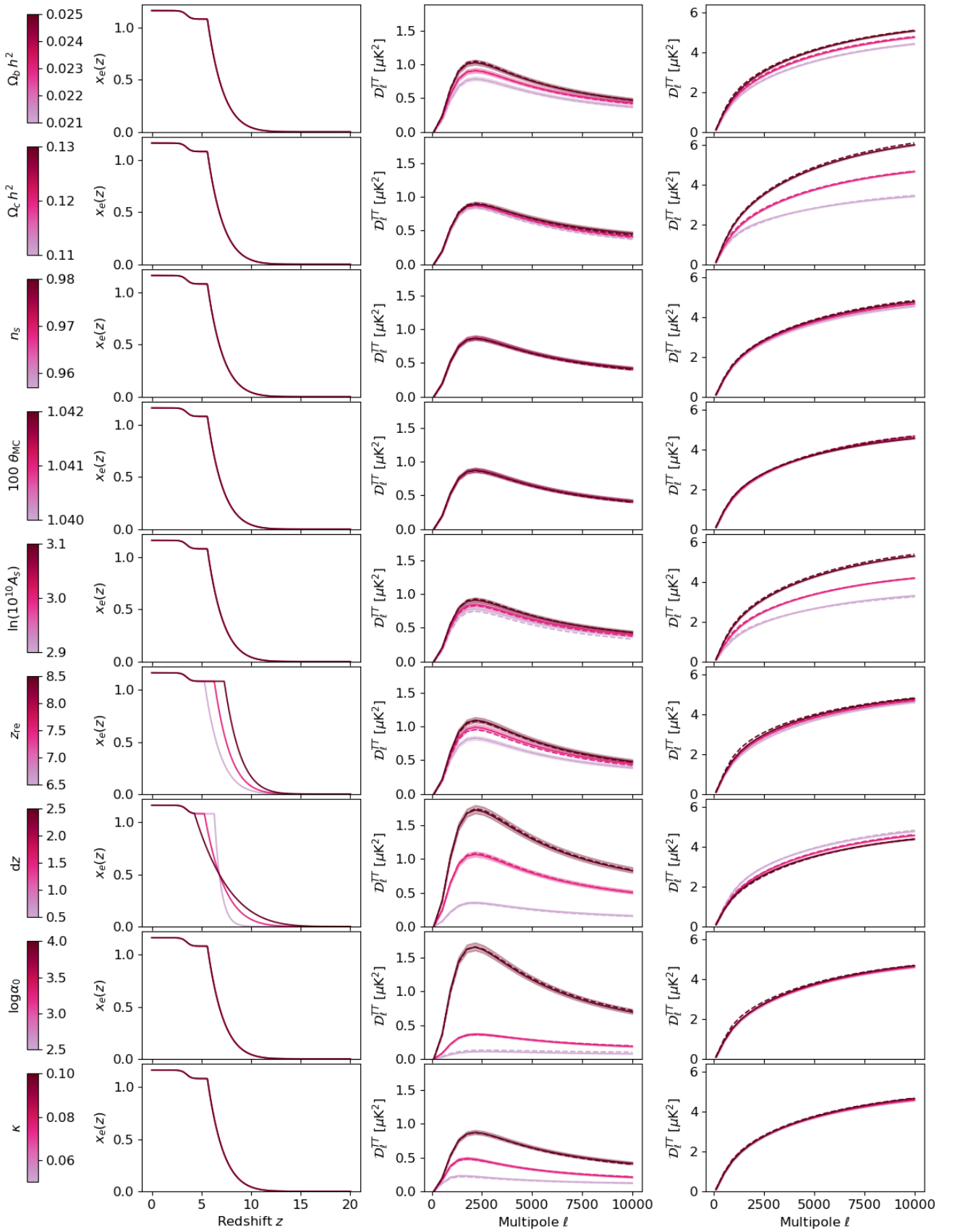


**Fig. A.1.** Patchy (upper panel) and late-time (low panel) kSZ spectra recovered by our random forest regressors, compared to true values. The lower panels show the 95% confidence intervals on the ratio of the two and the 95% confidence intervals on the absolute difference between the two, that is the absolute error.

late-time signal will be integrated over a larger redshift range. Additionally, the strong dependency of both kSZ components to the baryon and cold dark matter densities is related to the large contribution of the matter power spectrum in the derivation of their spectra. These dependencies are illustrated on Fig. A.2 and we refer the interested reader to Chen et al. (2022) for a detailed discussion of the dependency of the shape and amplitude of the patchy kSZ on various reionisation parameters. We emphasise the fact that the amplitude of the patchy and late-time kSZ spectra increases slowly with  $z_{re}$ , which will limit the constraining power of kSZ measurements on this parameter.

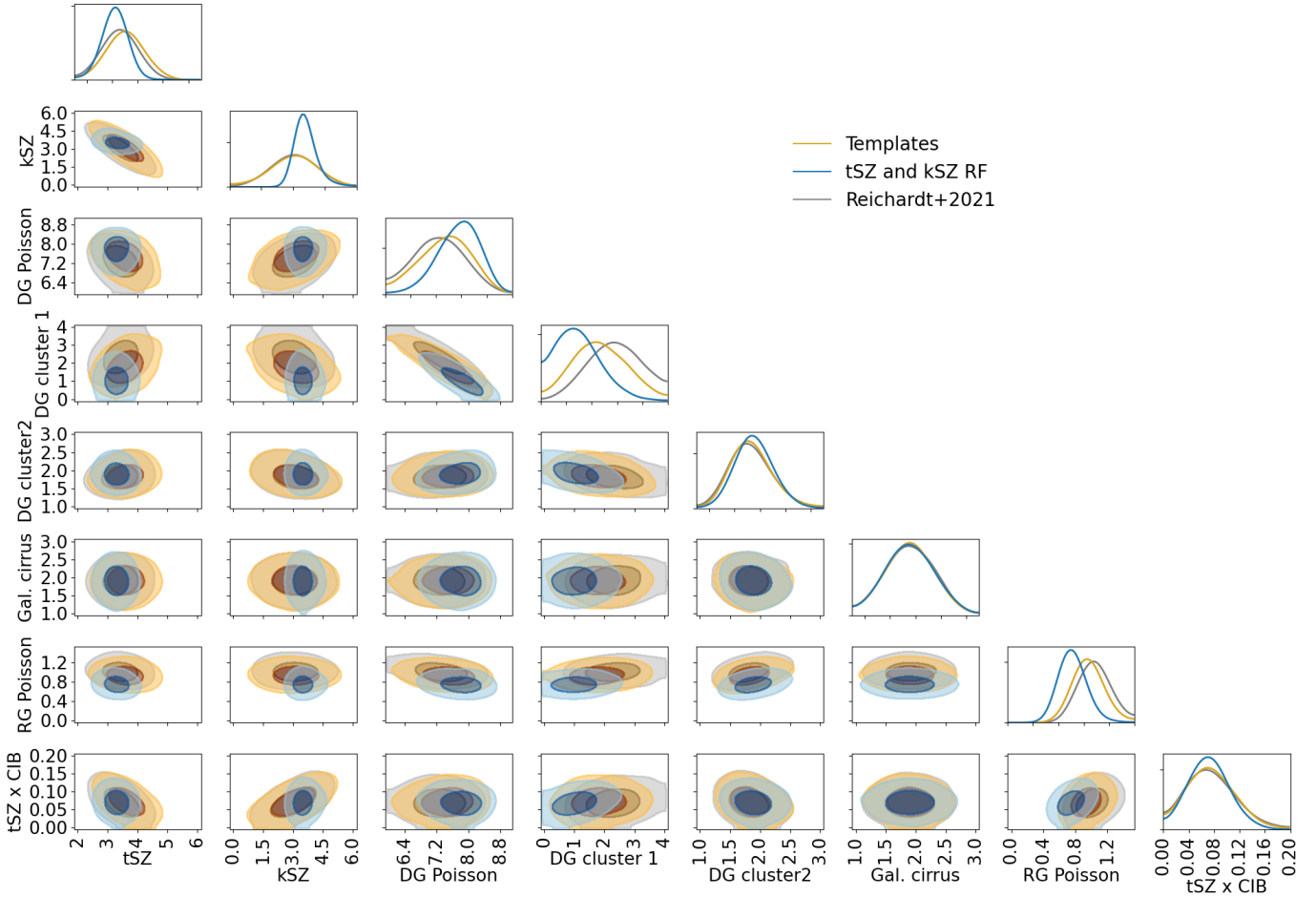
We show in Fig. A.1 some of the kSZ spectra recovered by the random forests after training, using parameters in the testing set – that is, parameter sets that the RF has not been explicitly trained on. We see that, overall, the late-time regressor performs better than the patchy one. They obtain, respectively, scores of 0.993 and 0.998 for a maximum of 1.0, corresponding to the coefficient of determination  $R^2$  of the prediction. This is likely because, as can be seen on the figure, there is much more variance between the patchy spectra than between the late-time spectra, and in particular there is more variance in their shape. However, in both cases, reconstruction errors remain within 5%

<sup>13</sup> Given the smoothness of the evolution of the kSZ spectra with  $\ell$ , we find this number of bins to be sufficient.



**Fig. A.2.** Dependence of the reionisation history (left panels), patchy (middle panels) and late-time (right panels) kSZ angular power spectra on the cosmological parameters the random forests were trained on (rows). For each parameter, three values are tested and we compare the spectrum obtained with the exact computation (dashed line) from the one inferred by the RF (solid line). The shaded area corresponds to the 68% confidence interval corresponding to the RF reconstruction errors (identical as Fig. A.1).





**Fig. A.3.** Posterior probability distributions of nuisance parameters obtained with SPT high- $\ell$  data (R21) using either templates (in yellow) or analytic derivation (in blue) for the tSZ and the kSZ power and a physical parameterisation of reionisation. Inferences are compared to the initial analysis, with templates and fixed reionisation (in grey). The CIBxtSZ amplitude is given in terms of its absolute value.

of the true value (1% for late-time), and are recovered with an absolute error  $<0.02 \mu\text{K}^2$ , which is about ten times smaller than the precision of the constraints obtained with the data (see Table 3). These errors are constant on the range of multipoles covered by the SPT data. To optimise the quality of the prediction, we have tested different values of hyper-parameters used by the RF, and we choose 40 for the number of trees and a depth of 20. These values maximise the score, whilst limiting the time and memory imprints of the reconstruction.

## Appendix B: Full parameter distributions

We present in this appendix additional posterior distributions from the different analyses carried out in this work, in order to support our arguments.

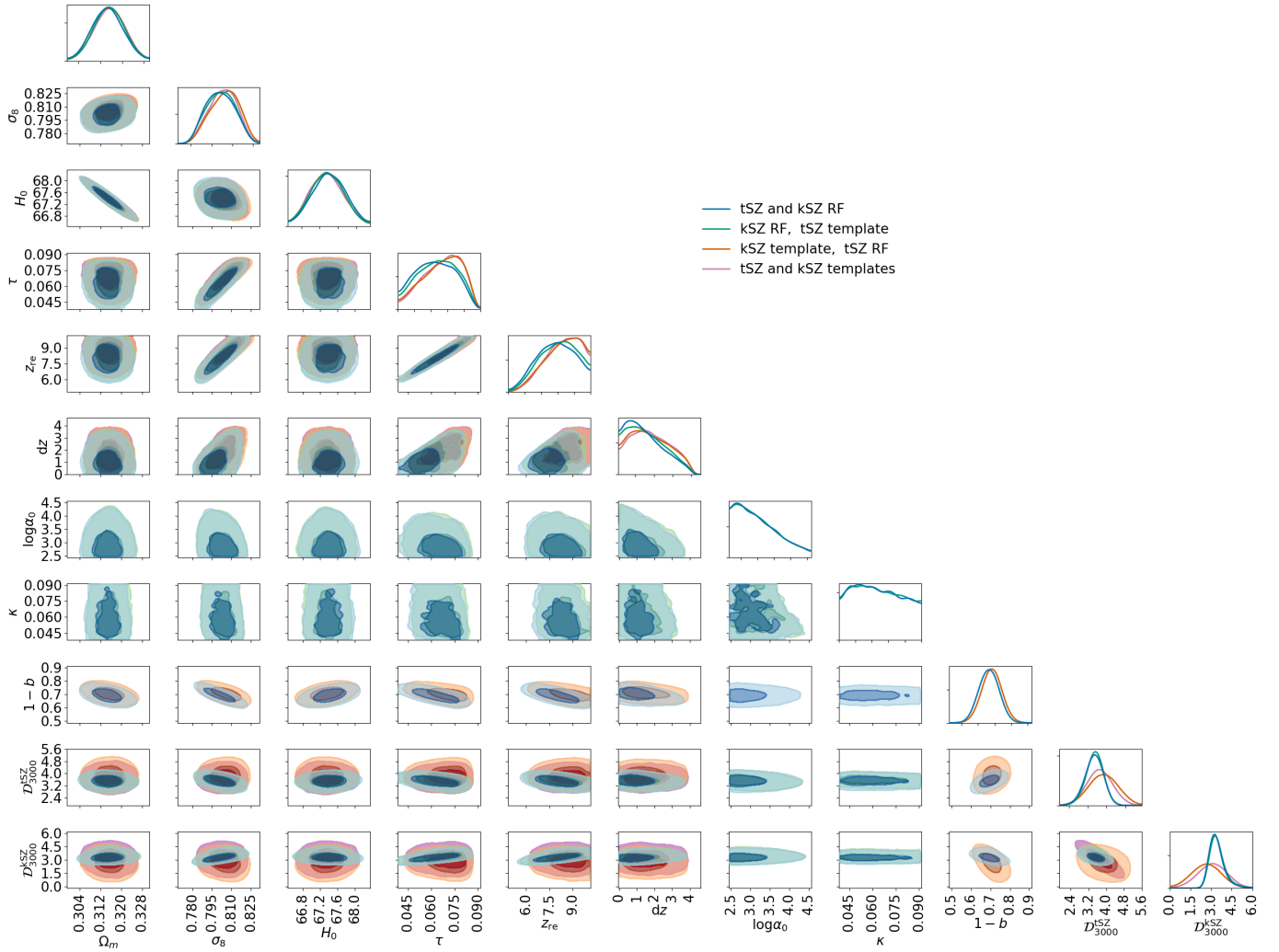
### B.1. Fixed cosmology

We present in this appendix the posterior distributions of the nuisance parameters – that is, of the amplitudes at  $\ell = 3000$  of the various foregrounds considered, for the run fitting the SPT data with fixed cosmology. Results are compared in Fig. A.3 between the results of R21, using templates for both SZ spectra; a similar analysis but with the reionisation model of Eq. (3); or with RF-inferred SZ spectra (see above). The constraints obtained with the cosmology-dependent SZ spectra are fully consistent

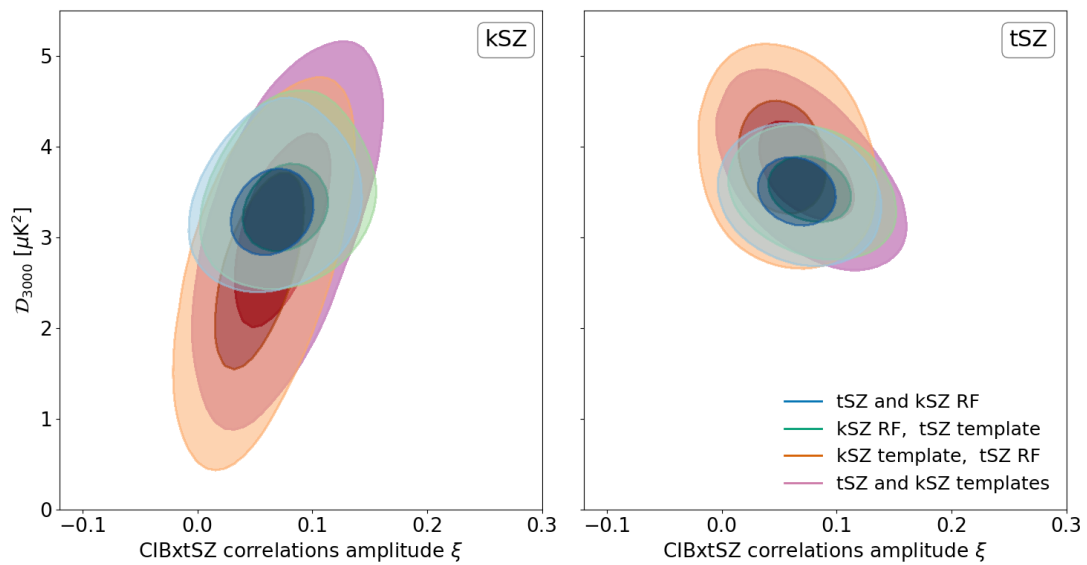
with previous results, although tighter. For example, the error bars on the contribution from Poisson-distributed dusty galaxies are reduced by 15%. Overall, the contribution of dusty and radio galaxies to the observed temperature power spectrum is smaller on all scales. Regarding the tSZxCIB correlations, the constraints on their amplitude at  $\ell = 3000$ ,  $\xi$ , obtained with the RF are fully consistent with previous results – that is, the data favours positive correlations at the  $2\sigma$  level, but the new model breaks the degeneracy between  $\xi$  and the SZ amplitudes.

### B.2. Varying cosmology

In Fig. B.1, we present the posterior distribution of the main cosmological and nuisance parameters obtained in Sec. 4.2. Depending on the nature of the analysis (that is, if templates or RF are used), these parameters can be sampled or derived. The cosmological information included in the SZ spectra when they are inferred by the RF at each iteration of the chain favours later and quicker reionisation scenarios although the constraints remain fully consistent with the results obtained with templates. The anti-correlation between  $\sigma_8$  and  $1 - b$ , as well as the correlation between  $1 - b$  and the tSZ amplitude, already observed in Paper I, is not broken by the addition of the cosmology-dependent kSZ model. Finally, the kSZ shape parameters,  $\log \alpha_0$  and  $\kappa$ , cannot be constrained precisely. Only an upper limit  $\log[\alpha_0/\text{Mpc}^3] < 4.1$  (95% confidence) can be placed on the



**Fig. B.1.** Posterior distributions of the different parameters sampled in the run with free cosmological and reionisation parameters using RF-predicted kSZ and tSZ spectra.



**Fig. B.2.** Joint posterior distributions of the kSZ and tSZ amplitudes at  $\ell = 3000$  with the amplitude of the tSZxCIB correlations at the same multipole,  $\xi$ . In all the results presented, the cosmology is free to vary, and different methods are used to model the contribution of the SZ effects to the observed small-scale temperature power spectrum of the CMB.

amplitude of the large-scale amplitude of the free electrons density fluctuations at high redshift. The interpretation of such an upper limit, in terms of high-redshift galaxy properties, will be the focus of future work.

We show in Fig. B.2 the joint posterior distributions of the amount of CIBxtSZ correlations,  $\xi$ , with the amplitudes of the kSZ and tSZ power spectra at  $\ell = 3000$ , for the four cases aforementioned. We see that, despite being fully consistent with the results obtained with templates, using the cosmology-dependent

SZ spectra allows to break the degeneracy between  $\xi$  and the SZ amplitudes. Replacing the tSZ template by the RF-inferred spectrum in the analysis allows to break the degeneracy between  $\xi$  and the tSZ amplitude: the correlation coefficient between  $\mathcal{D}_{3000}^{\text{tSZ}}$  and  $\xi$  drops from  $-0.6$  when the model used for the tSZ is a template, to  $-0.2$  when the cosmology-dependent tSZ power is used. Similarly, exploiting the cosmological information included in the kSZ power breaks the degeneracy between  $\mathcal{D}_{3000}^{\text{kSZ}}$  and  $\xi$ , moving from  $0.6$  to  $0.2$  when the kSZ template is replaced by the RF.