

An Optimal Approach for Measuring the Integrated Sachs Wolfe Effect

F.-X. Dupé¹ *, A. Rassat¹ **, J.-L. Starck¹, and M. J. Fadili²

¹ Laboratoire AIM, UMR CEA-CNRS-Paris 7, Irfu, SAp/SEDI, Service d'Astrophysique, CEA Saclay, F-91191 GIF-SUR-YVETTE CEDEX, France.

² GREYC UMR CNRS 6072, Université de Caen Basse-Normandie/ENSICAEN, 6 bdv Maréchal Juin, 14050 Caen, France.

Preprint online version: March 29, 2011

ABSTRACT

Context. One of the main challenges of modern cosmology is to understand the nature of the mysterious dark energy which causes the cosmic acceleration. The Integrated Sachs-Wolfe (ISW) effect is sensitive to dark energy **and if detected in a universe where modified gravity and curvature are excluded, presents an independent signature of dark energy**. The ISW effect occurs on large scales, where cosmic variance is high and where there are large amounts of missing data in the CMB and large scale structure maps due to Galactic confusion. Moreover, existing methods in the literature often make strong assumptions about the statistics of the underlying fields or estimators. Together these effects can severely limit signal extraction.

Aims. We want to define an optimal statistical method for detecting the ISW effect, which can handle large areas of missing data and minimise the number of underlying assumptions made about the data and estimators.

Methods. We first review current detections (and non-detections) of the ISW effect, comparing statistical subtleties between existing methods, and identifying several limitations. We propose a novel method to detect and measure the ISW signal. This method assumes only that the primordial CMB field is Gaussian. It is based on a sparse inpainting method to reconstruct missing data and uses a bootstrap technique to avoid assumptions about the statistics of the estimator. It is a complete method, which uses three complementary statistical methods.

Results. We apply our method to Euclid-like simulations and show we can expect a $\sim 7\sigma$ model-independent detection of the ISW signal with WMAP7-like data, even when considering missing data. Other tests return $\sim 4.7\sigma$ detection levels for a Euclid-like survey. We find detection levels are independent from whether the galaxy field is normally or lognormally distributed. We apply our method to the 2 Micron All Sky Survey (2MASS) and WMAP7 CMB data and find detections in the $1.1 - 2.0\sigma$ range, as expected from our simulations. As a by-product, we have also reconstructed the full-sky temperature ISW field due to 2MASS data.

Conclusions. We have presented a novel technique, based on sparse inpainting and bootstrapping, which accurately detects and reconstructs the ISW effect.

Key words. ISW, inpainting, signal detection, hypothesis test, bootstrap

1. Introduction

The recent abundance of cosmological data in the last few decades (for an example of the most recent results see Komatsu et al. 2009; Percival et al. 2007 a; Schrabback et al. 2010) has provided compelling evidence towards a standard concordance cosmology, in which the Universe is composed of approximately 4% baryons, 26% 'dark' matter and 70% 'dark' energy. One of the main challenges of modern cosmology is to understand the nature of the mysterious dark energy which drives the observed cosmic acceleration (Albrecht et al. 2006; Peacock et al. 2006).

The Integrated Sachs-Wolfe (ISW) (Sachs & Wolfe 1967) effect is a secondary anisotropy of the Cosmic Microwave Background (CMB), which arises because of the

variation with time of the cosmic gravitational potential between local observers and the surface of last scattering. The potential can be traced by Large Scale Structure (LSS) surveys (Crittenden & Turok 1996), and the ISW effect is therefore a probe which links the high redshift CMB with the low redshift matter distribution and can be detected by cross-correlating the two.

As a cosmological probe, the ISW effect has less statistical power than weak lensing or galaxy clustering (see for e.g., Refregier et al. 2010), but it is directly sensitive to dark energy, curvature or modified gravity (Kamionkowski & Spergel 1994; Kinkhabwala & Kamionkowski 1999; Carroll et al. 2005; Song et al. 2007), such that in universes where modified gravity and curvature are excluded, detection of the ISW signal provides a direct signature of dark energy. In more general universes, the ISW effect can be used to trace alternative models of gravity.

* francois-xavier.dupe@cea.fr

** anais.rassat@cea.fr

The CMB WMAP survey is already optimal for detecting the ISW signal (see Sections 2 and 3), and significance is not expected to increase with the arrival of Planck, unless the effect of the foreground Galactic mask can be reduced. The amplitude of the measured ISW signal should however depend strongly on the details of the local tracer of mass. Survey optimisations (Douspis et al. 2008) show that an ideal ISW survey requires the same configuration as surveys which are optimised for weak lensing or galaxy clustering - meaning that an optimal measure of the ISW signal will essentially come ‘for free’ with future planned weak lensing and galaxy clustering surveys (see for e.g. the Euclid survey Refregier et al. 2010). In the best scenario, a 4σ detection is expected (Douspis et al. 2008), and it has been shown that combined with weak lensing, galaxy correlation and other probes such as clusters, the ISW can be useful to break parameter degeneracies (Refregier et al. 2010), making it a promising probe.

Initial attempts to detect the ISW effect with COBE as the CMB tracer were fruitless (Boughn & Crittenden 2002), but since the arrival of WMAP data, tens of positive detections have been made, with the highest significance reported for analyses using a tomographic combination of surveys (see Sections 2 and 3 for a detailed review of detections). However, several studies using the same tracer of LSS appear to have contradicting conclusions, some analyses do not find correlation where others do, and as statistical methods to analyse the data evolve, the significance of the ISW signal is **sometimes** reduced (see for e.g., Afshordi et al. 2004; Rassat et al. 2007; Francis & Peacock 2010b).

In Section 2, we describe the cause of the ISW effect and review current detections. In Section 3, we describe the methodology for detection and measuring the ISW signal, and review a large proportion of reported detections in the literature, as well as their advantages and disadvantages. Having identified the main issues with current methods, we propose a new and complete method in Section 4, which capitalises on the fact that different statistical methods are complementary and uses sparse inpainting to solve the issue of missing data and a bootstrapping technique to measure the estimator’s probability distribution function (PDF). In Section 5, we validate our new method using simulations for **2MASS and Euclid-like surveys**. In Section 6, we apply our new method to WMAP 7 and the 2MASS survey. In Section 7, we present our conclusions.

2. The Integrated Sachs-Wolfe Effect

2.1. Origin of the Integrated Sachs-Wolfe Effect

General relativity predicts that the wavelength of electromagnetic radiation is sensitive to gravitational potentials, an effect which is called *gravitational redshift*. Photons travelling from the surface of last scattering will necessarily travel through the gravitational potential of Large Scale Structure (LSS) on their way to the observer; these will be blue-shifted as they enter the potential well and red-shifted as they exit the potential. These shifts will accumulate along the line of sight of the observer. The total shift in wavelength will translate into a change in the measured temperature-temperature anisotropy of the CMB, and can

be calculated by:

$$\left(\frac{\Delta T}{T}\right)_{ISW} = -2 \int_{\eta_L}^{\eta_0} \Phi'((\eta_0 - \eta)\hat{\mathbf{n}}, \eta) d\eta, \quad (1)$$

where T is the temperature of the CMB, η is the conformal time, defined by $d\eta = \frac{dt}{a(t)}$ and η_0 and η_L represent the conformal times today and at the surface of last scattering respectively. The unit vector $\hat{\mathbf{n}}$ is along the line of sight and the gravitational potential $\Phi(\mathbf{x}, \eta)$ depends on position and time. The integral depends on the rate of change of the potential $\Phi' = d\Phi/d\eta$.

In a universe with no dark energy or curvature, the cosmic (linear) gravitational potential does not vary with time, so that such a blue- and red-shift will always cancel out, because $\Phi' = 0$ and there will be no net effect on the wavelength of the photon.

However, in the presence of dark energy or curvature (Sachs & Wolfe 1967; Kamionkowski & Spergel 1994; Kinkhabwala & Kamionkowski 1999), the right hand side of Equation 1 will be non-null as the cosmic potential will change with time (see for e.g., Dodelson 2003), resulting in a secondary anisotropy in the CMB temperature field.

2.2. Detection of the ISW signal

The ISW effect leads to a linear scale secondary anisotropy in the temperature field of the CMB, and will thus affect the CMB temperature power spectrum at large scales. Due to the primordial anisotropies and cosmic variance on large scales, the ISW signal is difficult to detect directly in the temperature map of the CMB, but Crittenden & Turok (1996) showed it could be detected through cross-correlation of the CMB with a local tracer of mass.

The first attempt to detect the ISW effect (Boughn & Crittenden 2002) involved correlating the Cosmic Microwave Background explorer data (Bennett et al. 1990, COBE) with XRB (Boldt 1987) and NVSS data (Condon et al. 1998). This analysis did not find a significant correlation between the local tracers of mass and the CMB. Since the release of data from the Wilkinson Microwave Anisotropy Probe (Spergel et al. 2003, WMAP) over 20 studies (see Table 1) have investigated cross-correlations between the different years of WMAP data and local tracers selected using various wavelengths: X-ray (Boldt 1987, XRB survey); optical (Agüeros et al. 2006; Adelman-McCarthy et al. 2008, SDSS galaxies), (Anderson et al. 2001, SDSS QSOs), (Doroshkevich et al. 2004, SDSS LRGs), (Maddox et al. 1990, APM); near infrared (Jarrett et al. 2000, 2MASS); radio (Condon et al. 1998, NVSS).

The full sky WMAP data have sufficient resolution on large scales that the measure of the ISW signal is cosmic variance limited. The best LSS probe of the ISW effect should include maximum sky coverage and full redshift coverage of the dark energy dominated era (Douspis et al. 2008). No such survey exists yet, so there is room for improvement on the ISW detection as larger and larger LSS surveys arise. For this reason, when we review the current ISW detections, we classify them according to their tracer of LSS, and not the CMB map used.

The measure of the ISW signal can be done in various statistical spaces; we classify detections in Table 1 into three measurement ‘domains’: D1 corresponds to spherical harmonic space; D2 to configuration space and D3 to

Author	CMB	LSS Tracer	Wavelength	Method	Claimed Detection
Boughn & Crittenden (2002)	COBE	XRB	Xray	D2	No
Giannantonio et al. (2008)	W3			D2	2.7σ
Boughn & Crittenden (2004, 2005)	W1	XRB/NVSS	Xray/Radio	D2	‘tentative’ (2-3 σ)
Fosalba et al. (2003)	W1	SDSS DR1		D2	2σ (low z) 3.6σ (high z)
Cabré et al. (2006)	W3	SDSS DR4	Optical	D2	$> 2\sigma$
Giannantonio et al. (2008)	W3	SDSS DR6		D2	2.2σ
Sawangwit et al. (2010)	W5	SDSS DR5		D2	‘marginal’
López-Corrodoira et al. (2010)	W5	SDSS DR7		D2	‘No detection’
Giannantonio et al. (2006)	W3	SDSS Quasars	Optical	D2	2σ
Giannantonio et al. (2008)	W3	SDSS Quasars		D2	2.5σ
Xia et al. (2009)	W5	SDSS Quasars		D2	2.7σ
Scranton et al. (2003)	W1			D2	$> 2\sigma$
Padmanabhan et al. (2005)	W1			D1	2.5σ
Granett et al. (2009)	W3	SDSS LRG	Optical	D1	2σ
Giannantonio et al. (2008)	W3			D2	2.2σ
Sawangwit et al. (2010)	W5	SDSS LRG, 2SLAQ		D2	‘marginal’
Sawangwit et al. (2010)	W5	AAOmega LRG		D2	Null
Fosalba & Gaztañaga (2004)	W1	APM	Optical	D2	2.5σ
Afshordi et al. (2004)	W1			D1	2.5σ
Rassat et al. (2007)	W3	2MASS	NIR	D1	2σ
Giannantonio et al. (2008)	W3			D2	0.5σ
Francis & Peacock (2010b)	W3			D1	‘weak’
Boughn & Crittenden (2002)	COBE			D2	No
Nolta et al. (2004)	W1			D2	2.2σ
Pietrobon et al. (2006)	W3	NVSS	Radio	D3	$> 4\sigma$
Vielva et al. (2006)	W3			D3	3.3σ
McEwen et al. (2007)	W3			D3	$> 2.5\sigma$
Raccanelli et al. (2008)	W3			D2	2.7σ
McEwen et al. (2008)	W3			D3	$\sim 4\sigma$
Giannantonio et al. (2008)	W3			D2	3.3σ
Hernández-Monteagudo (2009)	W3			D1	$< 2\sigma$
Sawangwit et al. (2010)	W5			D2	‘marginal’ ($\sim 2\sigma$)
Corasaniti et al. (2005)	W1			D2	$> 2\sigma$
Gaztañaga et al. (2006)	W1			D2	2σ
Ho et al. (2008)	W3	Combination	Combination	D1	3.7σ
Giannantonio et al. (2008)	W3			D2	4.5σ

Table 1. Meta-analysis of ISW detections to date and their reported statistical significance. The ‘Method’ describes the space in which the power spectrum analysis is done (configuration, spherical harmonic, etc ...), not the method for measuring the significance level of the detection (this is described in Section 3). **D1** corresponds to spherical harmonic space, **D2** to configuration space, **D3** to wavelet space. The highest detections are made in wavelet space. Regarding the survey used, the highest detections are made using NVSS (though weak and marginal detections using NVSS are also reported) or using combinations of LSS surveys as the matter tracer.

wavelet space. (In Section 3, we review the different methods for quantifying the statistical significance of each measurement).

There are only two analyses which use COBE as CMB data (with XRB and NVSS data, Boughn & Crittenden 2002), and both report null detections, which can reasonably be due to the low angular resolution of COBE even at large scales. The rest are done correlating WMAP data from years 1, 3 and 5 (respectively ‘W1’, ‘W3’ and ‘W5’ in table 1).

Most ISW detections reported in Table 1 are relatively ‘weak’ ($< 3\sigma$) and this is expected from theory for a concordance cosmology. Higher detections are reported for the NVSS survey (Pietrobon et al. 2006; McEwen et al. 2008; Giannantonio et al. 2008), though weak and marginal de-

tectations using NVSS data are also reported (Hernández-Monteagudo 2009; Sawangwit et al. 2010). High detections are often made using a wavelet analysis (Pietrobon et al. 2006; McEwen et al. 2008), though a similar study by the same authors using the same data but a different analysis method finds a weaker signal (McEwen et al. 2007). The highest detection is reported using a tomographic combination of all surveys (XRB, SDSS galaxies, SDSS QSOs, 2MASS and NVSS, Giannantonio et al. 2008), as expected given the larger redshift coverage of the analysis.

Several analyses have been revisited to seek confirmation of previous detections. In some cases, results are very similar (Padmanabhan et al. (2005); Granett et al. (2009); Giannantonio et al. (2008), for SDSS LRGs; Giannantonio et al. (2006, 2008) for SDSS Quasars; Afshordi et al. (2004);

Rassat et al. (2007), for 2MASS), but in some cases they are controversially different (for e.g. Pietrobon et al. (2006) and Sawangwit et al. (2010), for NVSS or Afshordi et al. (2004) and Giannantonio et al. (2008), for 2MASS).

We also notice that as certain surveys are revisited, there is a trend for the statistical significance to be reduced: for e.g., detections from 2MASS decrease from a 2.5σ detection (Afshordi et al. 2004), to 2σ (Rassat et al. 2007), to 0.5σ (Giannantonio et al. 2008) to ‘weak’ (Francis & Peacock 2010b). Detections using SDSS LRGs decrease from 2.5σ (Padmanabhan et al. 2005), to $2 - 2.2\sigma$ (Granett et al. 2009; Giannantonio et al. 2008), to ‘marginal’ (Sawangwit et al. 2010). Furthermore, there tends to be a ‘sociological bias’ in the interpretation of the confidence on the signal detection. The first detections interpret a $2 - 3\sigma$ detection as ‘tentative’ (Boughn & Crittenden 2004, 2005), while further studies with similar detection level report ‘independent evidence of dark energy’ (Afshordi et al. 2004; Gaztañaga et al. 2006).

3. Methodology for Detecting the ISW Effect

In this paper, we are interested in qualifying the differences between different statistical methods which exist in the literature, and comparing them with a new method we present in Section 4. By statistical method, we mean the method which is used to quantify the significance of a signal, not the space in which the signal is measured. Therefore, and without loss of generality, the review presented in Section 3.2 summarises methods using spherical harmonics. We compare the pros and cons of each method in Section 3.3. We begin by describing how the ISW signal can be measured in spherical harmonics in Section 3.1

3.1. ISW Signal in Spherical Harmonics

In general, any field can be decomposed by a series of functions which form an orthonormal set, as do the spherical harmonic functions $Y_{\ell m}(\theta, \phi)$. Therefore a projected galaxy overdensity (δ_g) or temperature anisotropy (δ_T) field $\delta_X(\theta, \phi)$, where $X = g, T$, can be decomposed into:

$$\delta_X(\theta, \phi) = \sum_{\ell, m} a_{\ell m}^X Y_{\ell m}(\theta, \phi), \quad (2)$$

where $a_{\ell m}^X$ are the spherical harmonic coefficients of the field. The 2-point galaxy-temperature cross-correlation function can then be written:

$$C_{gT}(\ell) = \frac{1}{(2\ell + 1)} \sum_m \mathcal{R}e [a_{\ell m}^g (a_{\ell m}^T)^*], \quad (3)$$

where taking the real part of the product ensures that $C_{gT}(\ell) = C_{Tg}(\ell)$.

The theory for the angular cross-correlation function is given by:

$$C_{gT}(\ell) = 4\pi b_g \int dk \frac{\Delta^2(k)}{k} W_g(k) W_T(k), \quad (4)$$

where

$$W_g(k) = \int dr \Theta(r) j_\ell(kr) D(z), \quad (5)$$

$$W_T(k) = -\frac{3\Omega_{m,0} H_0^2}{k^2 c^3} \int_0^{z_L} dr j_\ell(kr) H(z) D(z) (f - 1), \quad (6)$$

$$\Delta^2(k) = \frac{4\pi}{(2\pi)^3} k^3 P(k), \quad (7)$$

$$\Theta(r) = \frac{r^2 n(r)}{\int dr r^2 n(r)}. \quad (8)$$

In these equations, r represents the co-moving distance, z_L the redshift at the last scattering surface, k the Fourier mode wavenumber and quantities which depend on the redshift z have an intrinsic dependence on r : $H(z) = H(z(r))$. The function f is the linear growth factor given by $f = \frac{d \ln D(z)}{d \ln a(z)}$, where $D(z)$ is the linear growth which measures the growth of structure. The cross-correlation function depends on the survey selection function given by $n(r)$, in units of galaxies per unit volume. The quantities $\Omega_{m,0}$ and H_0 are the values of the matter density and the Hubble parameter at $z = 0$. Units are chosen so that the quantity $C(\ell)$ is unitless.

In the case where both the temperature and the galaxy fields behave as Gaussian random fields, then the covariance on the ISW signal can be calculated by:

$$\langle |C_{gT}|^2 \rangle = \frac{1}{f_{sky} (2\ell + 1)} [C_{gT}^2 + (C_{gg} + \mathcal{N}_g)(C_{TT} + \mathcal{N}_T)], \quad (9)$$

where C_{TT} is the temperature-temperature power spectrum, \mathcal{N}_g and \mathcal{N}_T are the noise of the galaxy and temperature fields respectively. The galaxy auto-correlation function can be calculated theoretically in linear theory by:

$$C_{gg}(\ell) = 4\pi b_g^2 \int dk \frac{\Delta^2(k)}{k} [W_g(k)]^2, \quad (10)$$

There are many difficulties in measuring the ISW effect, the first being the intrinsic weakness of the signal. To add to this, an unknown galaxy bias scales linearly with the ISW cross-correlation signal (see Equation 4), which is therefore strongly degenerate with cosmology. Galactic foregrounds in both the CMB and the LSS maps also mask crucial large scale data and can introduce spurious correlations. Any method claiming to detect the ISW effect should be as thorough as possible in accounting for missing data, and where possible the reported detection level should be independent of an assumed cosmology.

3.2. Review on Current Tools for ISW Detection

In the literature there are two quantities which can be used to measure and detect the ISW signal, which we review in this section. Without loss of generality, we present these methods in spherical harmonic space. The first method measures the observed cross-correlation spectra (‘Spectra’ method: see section 3.2.3), whilst the second directly compares temperature fields (‘Fields’ method: see section 3.2.4). These two approaches differ by the quantity they measure to infer a detection. For each method (fields vs. spectra), it is possible to use different statistical methods to infer detection which we describe below. We review each existing method below and summarise the pros and cons of both of these classes as well as the statistical models in Table 2.

3.2.1. Note on the confidence score

Before reviewing the ISW detection methods, we would like to clarify the definition of confidence

Measured Quantity	Advantage	Disadvantage
Spectra	Methods exist for calculation with missing data. Can introduce tomography easily.	Assumes $C(\ell)$'s or estimator are Gaussian. Most methods requires estimation of the covariance matrix.
Fields	No assumption about galaxy/matter density field	Missing data is an ill-posed problem.
Statistical Method	Advantage	Disadvantage
Simple correlation (Spectra / Fields)	Independent of cosmology.	Measure of significance assumes estimator is Gaussian.
Amplitude estimation (Spectra / Fields)	Validates signal and model simultaneously.	Detection depends on cosmology/model. Measure of significance assumes estimator is Gaussian. Assumes underlying theory (e.g. Λ CDM) is correct.
χ^2 (Spectra / Fields)	Validates signal and model simultaneously.	Assumes $C(\ell)$'s are Gaussian (Spectra). Only gives confidence of rejecting null hypothesis. Assumes underlying theory (e.g. Λ CDM) is correct.
Model comparison (Spectra)	Asks a different question than other tests.	ISW signal usually too weak to be detected this way. Assumes underlying theory (e.g. Λ CDM) is correct.

Table 2. TOP: Review of advantages and disadvantages of measuring spectra vs. fields in order to infer an ISW detection. BOTTOM: Review of statistical methods *existing in the literature* and their respective advantages and disadvantages.

scores from a statistical point of view. The confidence of a null hypothesis test can be interpreted as the distance from the data to the null hypothesis (commonly named H_0). For example, let ρ be a variable of interest (e.g. correlation coefficient, amplitude). The confidence score σ for the hypothesis test H_0 (i.e., $\rho = 0$) against H_1 (i.e., $\rho \neq 0$) is directly computed using the formula $\tau = \rho/\sigma(\rho)$, where $\sigma(\rho)$ is the standart deviation. But this only true when 1) ρ is Gaussian, 2) $\sigma(\rho)$ is computed independently of the observation and 3) considering a symmetric test. Then, this method does not stand for the general case and as the correlation coefficient considered here must be positive, an asymmetric test (one-sided test) would also be more appropriate here.

Remember that a confidence score is directly linked to the deviation from the H_0 hypothesis through the p -value which is a probability, so the σ -score is always positive. Then the p -value p of one hypothesis test is computed using the probability density function (PDF) of the test distribution: $p = 1 - \int_{-\infty}^{\tau} P(x|H_0)dx = \int_{\tau}^{+\infty} P(x|H_0)dx$ (for a classical one-sided test). Remark that in that case, if the H_0 is true then $\rho \approx 0$ (i.e. in the middle of the test distribution) and the p -value p will be around 0.5 which correspond to a confidence of 0.67σ .

3.2.2. Note on the application spaces

All the methods that will be described in the next sections can be performed in different domains. While some spaces may be more appropriate than others for a specific task, difficulties may also arise because of the properties of the space. For example: for the ‘spectra’ method in configuration space, the two main difficulties are missing data and the estimation of the covariance matrix (see e.g. Hernández-Monteaugudo 2008). In this case, the covariance matrix can be estimated using Monte Carlo methods (see Cabré et al. 2007). In spherical harmonic space, missing data induce mode correlations which can be removed by using an appro-

priate framework for calculation of the $C(\ell)$'s (e.g. Hivon et al. 2002). In harmonic space (when missing data is accounted for) the covariance matrix is diagonal and thus easily invertible (see Section 3.2.4).

3.2.3. Cross-power spectra comparison

The most popular method consists in using the cross-correlation function (Equation 4, in spherical harmonic space) to measure the presence of the ISW signal, however this approach has recently been challenged by López-Corredoira et al. (2010), because of its high sensitivity to noise and fluctuations due to cosmic variance.

One of the subtleties of the cross-correlation function method is the evaluation of the covariance matrix C_{covar} and its inverse. This matrix can be estimated using the MC1 or MC2 methods of Cabré et al. (2007), in which case the test is strongly dependent on the quality of the simulations. Secondly, missing data will require extra care when estimating the power spectra, this can be tackled by using MASTER (Monte Carlo Apodized Spherical Transform Estimator) or QML (Quadratic Maximum Likelihood) methods (Hivon et al. 2002; Efstathiou 2004; Munshi et al. 2009).

The spectra measurement can be used with one of four different statistical methods. The advantages and disadvantages of each method are summarised below and in Table 2. The first aims to detect a correlation between two signals, i.e. we test if the cross-power spectra is null or not. The second fits a (model-dependent) template to the measured cross-power spectra. The other two methods aim to validate a cosmological model as well as confirm the presence of a signal: the χ^2 test and the model comparison. We describe them below:

- **Simple correlation detection:** The simplest and the most widely used method for detecting a cross-correlation between two fields X and Y (here supposing that Y is correlated with X) (see e.g. Boughn & Crittenden 2002; Afshordi et al. 2004; Pietrobon et al. 2006; Sawangwit et al. 2010) is to measure the correla-

tion coefficient $\rho(X, Y)$, defined as:

$$\rho(X, Y) = \text{Cor}(X, Y) / \text{Cor}(X, X),$$

where, $\text{Cor}(X, Y) = \frac{1}{N_p} \sum_p \text{Re}[X^*(p)Y(p)]$, (11)

with p a position or scale parameter and N_p the number of considered positions or scales. There is a correlation between the two fields if $\rho(X, Y)$ is not null. The correlation coefficient is linked to the cross-power spectra in harmonic space:

$$\rho(g, T) = \left(\sum_{\ell} (2\ell + 1) C_{gT}(\ell) \right) / \left(\sum_{\ell} (2\ell + 1) C_{gg}(\ell) \right). \quad (12)$$

Thus the nullity of the coefficient implies the nullity of the cross-power spectra. A z-score can be performed in order to test this nullity,

$$K_0 = \rho / \sigma_{\rho}, \quad (13)$$

where the standard error of the correlation value, σ_{ρ} can be estimated using Monte Carlo simulations under a given cosmology. In the literature, most applications of this method assume that \mathbf{K}_0 follows a Gaussian distribution under the null hypothesis (i.e. no correlation), for example Vielva et al. (2006) and Giannantonio et al. (2008), which is not necessarily true.

The distribution of the K_0 test can be inferred if we assume that both fields X and Y , of the cross-correlation are Gaussian. In this case, the correlation coefficient distribution is the normally distributed but follows a normal product distribution, which is far from Gaussian. In the case where Y is a constant field, the correlation coefficient follows a normal distribution **and the distribution of the hypothesis test K_0 will depend on how the variance of the estimator σ_{ρ} is computed. If this last value is derived from the observation X , then K_0 follows a Student's t-distribution which converges to a Gaussian distribution only when ρ is high (by the central limit theorem). Else, if the variance σ_{ρ} is estimated independently from the data (through Monte-Carlo, for example) or known for a given cosmology, K_0 can be assumed to follow a Gaussian distribution. This means that K_0 is generally not Gaussian even if the correlation coefficient is Gaussian (see section 4).**

This method does not include knowledge of the underlying ISW signal, nor of the galaxy field, though the error bars can be estimated from Monte Carlo simulations which include cosmological information.

– **Amplitude estimation (or template matching):**

The principle of the amplitude estimation is to measure whether an observed signal corresponds to the signal predicted by a given cosmological model.

The estimator and its variance are given by (e.g. Ho et al. 2008; Giannantonio et al. 2008):

$$\hat{\lambda} = \frac{C_{gT}^{\text{Th}*} C_{\text{covar}}^{-1} C_{gT}^{\text{Obs}}}{C_{gT}^{\text{Th}*} C_{\text{covar}}^{-1} C_{gT}^{\text{Th}}}, \quad \sigma_{\hat{\lambda}} = \frac{1}{\sqrt{C_{gT}^{\text{Th}*} C_{\text{covar}}^{-1} C_{gT}^{\text{Th}}}}, \quad (14)$$

where C_{gT}^{Th} is the theoretical cross-power spectrum, C_{gT}^{Obs} the estimated (observed) power spectrum and

C_{covar} the covariance matrix calculated by Equation 9. A z-score,

$$K_1 = \hat{\lambda} / \sigma_{\hat{\lambda}}, \quad (15)$$

is usually applied to test if the amplitude is null or not.

– **Goodness of fit, χ^2 test:**

The goodness of fit or χ^2 test is given by (e.g. Afshordi et al. 2004; Rassat et al. 2007):

$$K_2 = (C_{gT}^{\text{Th}} - C_{gT}^{\text{Obs}})^* C_{\text{covar}}^{-1} (C_{gT}^{\text{Th}} - C_{gT}^{\text{Obs}}), \quad (16)$$

where K_2 follows a χ^2 distribution with number of degrees of freedom (*d.o.f*) depending on the input data. This tests the correspondance of the data with a given cosmological model, but does not infer if the tested model is in fact the best. Equation 16 also assumes that the $C(\ell)$'s are Gaussian variables. The value of χ^2 gives a idea on the probability of rejecting the model, but cannot be compared directly with the K_2 value for the null hypothesis without careful statistics (see next method on model comparison).

– **Model comparison:**

The model comparison method is based on the generalised likelihood ratio test and asks the question: ‘*Do the data prefer a given fiducial cosmological model over the null hypothesis ?*’. This question is important because it could be possible to use the previous χ^2 test to detect an ISW signal - yet the data could still also be compatible with a null hypothesis (see for e.g., Afshordi et al. 2004; Rassat et al. 2007; Francis & Peacock 2010b). In this case it is important to perform a model comparison to find which model is preferred by the data. Two hypotheses are built:

- H_0 : “*there is no ISW signal*”, i.e. the cross-power spectra is null;
- H_1 : “*there is an ISW signal compatible with a fiducial cosmology*”, i.e. the cross-power spectra is close to an expected one.

One should then estimate:

$$K_3 = \Delta\chi^2 = (C_{gT}^{\text{Obs}})^* C_{\text{covar}}^{-1} (C_{gT}^{\text{Obs}} - (C_{gT}^{\text{Th}} - C_{gT}^{\text{Obs}})^* C_{\text{covar}}^{-1} (C_{gT}^{\text{Th}} - C_{gT}^{\text{Obs}})), \quad (17)$$

where K_3 converges asymptotically to a χ^2 distribution. If the value is higher than a threshold (chosen for a required confidence level), the H_0 hypothesis is rejected. However, such method is difficult to use directly because of the small sample bias, K_3 is not likely to follow a χ^2 statistic. In the case of the ISW signal, the signal for a ‘standard’ fiducial cosmology (e.g., WMAP 7 cosmology) is so weak that it usually returns a lack of detection for current surveys - this may not be the case for future or tomographic surveys. Notice that this model comparison method can be seen as an improved version of the goodness of fit.

3.2.4. Field to field comparison

Instead of comparing the spectra, one can work directly with the temperature field to measure the presence of the ISW signal. The observable in this case is now the ISW temperature field (δ_{ISW}), rather than the cross-correlation power spectra $C_{gT}(\ell)$. The observed CMB temperature anisotropies δ_{OBS} can be described as:

$$\delta_{\text{OBS}} = \delta_T + \lambda\delta_{\text{ISW}} + \delta_{\text{other}} + \mathcal{N}, \quad (18)$$

where δ_{ISW} is the ISW field and λ its amplitude (normally near 1), δ_T the primordial CMB temperature field, δ_{other} represents fluctuations due to secondary anisotropies other than the ISW effect and \mathcal{N} represents noise. In the context of the ISW effect, which occurs only on large (linear) scales where noise and other secondary anisotropies are negligible, we have:

$$\delta_{\text{OBS}} \simeq \delta_T + \lambda \delta_{\text{ISW}} . \quad (19)$$

The main difference between the fields and spectra approach is that the fields method requires an estimation of the ISW temperature field (δ_{ISW}). There are several methods to calculate δ_{ISW} from a given matter overdensity map. **The most accurate way to reconstruct the ISW signal is to use information from the full 3-dimensional matter distribution, which in theory requires overlapping galaxy and weak lensing maps on large scales. This may be possible in the future with surveys like Euclid (Refregier et al. 2010). Assuming a simple bias relation, the matter field can also be estimated directly from galaxy surveys (see Granett et al. 2009, who did this for small patches on the sky). In the case where only the general redshift distribution of the galaxy survey is known, the ISW field δ_{ISW} can be approximated directly from the galaxy and temperature maps using (see Cabré et al. 2007):**

$$a_{\ell m}^{\text{ISW}} = \frac{C_{gT}(\ell)}{C_{gg}(\ell)} g_{\ell m} , \quad (20)$$

where $g_{\ell m}$ are the spherical harmonic coefficients of the galaxy map, and $a_{\ell m}^{\text{ISW}}$ the coefficients of the ISW temperature anisotropy map. Another approach is to reconstruct the ISW map using Equation 1 where Φ' is estimated using the Poisson Equation (Francis & Peacock 2010a).

Note that the potential which creates the ISW signal is first order in perturbation theory. However, even in the absence of non-Gaussianities in the primordial inflaton field, the potential ϕ will evolve with redshift, and can become lognormal on quasi-linear scales, where the potential is still decaying: this will produce an ISW temperature signal (which is positively correlated with the density field), which will approach Gaussianity on the largest scales but may contain some traces of non-Gaussianity on smaller scales (see Francis & Peacock 2010a,b). On smaller and non-linear scales, the Rees-Sciama effect will produce secondary temperature anisotropies which are negatively correlated with the density field (Schaefer et al. 2010; Cai et al. 2010).

As for the spectra approach, there are several statistical methods available to qualify detection:

– **Simple correlation detection:**

The simple correlation detection method described for the spectra comparison, can in fact also be considered as a field comparison. This is the only method which directly overlaps between both approaches.

– **Amplitude estimation (or template matching):**

Using the Gaussian framework, given an ISW field, the amplitude λ can be estimated with the corresponding maximum likelihood estimator (Hernández-Monteagudo 2008; Frommert et al. 2008; Granett et al.

2009):

$$\hat{\lambda} = \frac{\delta_{\text{ISW}}^* C_{TT}^{-1} \delta_{\text{OBS}}}{\delta_{\text{ISW}}^* C_{TT}^{-1} \delta_{\text{ISW}}} , \quad \sigma_{\hat{\lambda}} = \frac{1}{\sqrt{\delta_{\text{ISW}}^* C_{TT}^{-1} \delta_{\text{ISW}}}} . \quad (21)$$

A signal is present if $\hat{\lambda}$ is non null and a z-score,

$$K_4 = \hat{\lambda} / \sigma_{\hat{\lambda}} , \quad (22)$$

directly yields the confidence level in terms of σ . Equation 21 implicitly assumes that the primordial CMB field δ_T is a Gaussian random field (we discuss this further in section 4).

– **Goodness of fit, χ^2 test:**

The χ^2 goodness of fit with H_1 (see **Model comparison** in 3.2.3) yields:

$$K_5 = (\delta_{\text{ISW}} - \delta_{\text{OBS}})^* C_{TT}^{-1} (\delta_{\text{ISW}} - \delta_{\text{OBS}}) , \quad (23)$$

where K_5 is a χ^2 variable with number of *d.o.f* depending on the input data. In this case, the test only returns the confidence of rejecting the null hypothesis H_0 . As in the χ^2 test for the spectra, precaution must be taken when comparing χ^2 values for different models, by using an appropriate model comparison technique. We introduce this in section 4.

3.3. Pros and cons of each method

We have identified two main classes of methods: either using power spectra or fields to measure the ISW signal. For each approach one can choose amongst several statistical tools to measure the significance of a correlation or validate simultaneously a correlation and a model. The advantages and disadvantages of both approaches are summarised below and in the top part of Table 2.

One of the main advantages of using the field approach is that it assumes only that the primordial CMB field comes from a Gaussian random process, which is largely believed to be true. In the other approach, the spectra are assumed to be Gaussian, which is not the case. Several studies (Cole et al. 2001; Kayo et al. 2001; Wild et al. 2005) have also shown that the matter overdensity exhibits a lognormal behavior on large scale. The bias introduced has been shown to be small (Bernardeau et al. 2002; Hamimeche & Lewis 2008), however this approach is still theoretically ill-motivated. The main advantage of the spectra method is the relative ease when calculating the spectra from incomplete data sets, as tools are available for calculating the spectra (see e.g. Efstathiou 2004). In the field approach, managing missing data is an ill-posed problem.

We remind that a problem is defined as a well-posed problem (Hadamard 1902) if 1) a solution exists, 2) the solution is unique and 3) the solution depends continuously on the data (in some reasonable topology). Otherwise, the problem is defined as a ill-posed problem. With missing data, the second point cannot be verified. Reconstruction of the data also requires inversion of an operation (e.g., the mask), and is therefore an *ill-posed inverse problem*. Notice that most inverse problems are ill-posed (e.g. deconvolution).

4. A rigorous method for detecting the ISW effect

Having identified in the previous section the numerous methods used in the literature to detect and measure the ISW signal, as well as their relative advantages and disadvantages, we propose here a complete and rigorous method for detecting and quantifying the signal significance. We describe this method in detail below and it is summarised in Figure 2.

4.1. Motivation

If we consider the pros and cons of each detection method shown in Table 2 and in Section 3.3, we remark first that any method based on the comparison of spectra makes the demanding assumption that the $C(\ell)$'s be Gaussian, whereas methods based on field comparison require only the primordial CMB field to be Gaussian. **Instead the fields method assumes only that the primordial CMB is Gaussian, so we recommend this method be used for an ISW analysis.**

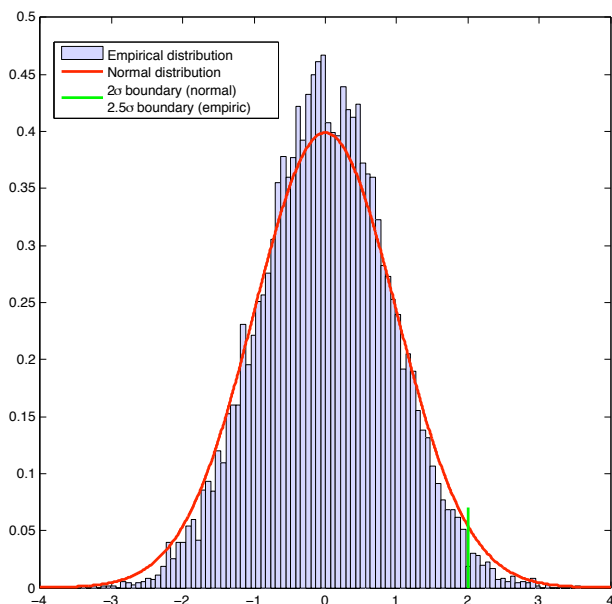


Fig. 1. Comparison of Gaussian PDF for estimator λ/σ_λ with its true estimated distribution. A 2σ significance using the Gaussian PDF corresponds in fact to a 2.5σ detection with the true distribution. (Calculation for 2MASS survey, see section 6).

Another key issue is the estimation of the signal significance. We find that most approaches in the literature assume that the probability distribution function (PDF) of the *estimator* is Gaussian. In Figure 1 we evaluate the estimator's PDF (under the null hypothesis) for the ISW signal due the 2MASS survey (see Section 6) using a Monte-Carlo method (purple bars) and compare it with a Gaussian PDF (red solid line). The distributions' tails differ which leads to a bias in the confidence level for positive data (i.e. data with an ISW signal). For the 2MASS survey, a 2σ detection with a Gaussian assumption (vertical solid/green line) corresponds in fact to a 2.5σ detection using the true underlying PDF. In this case, the signal amplitude is underestimated with the Gaussian assumption. Such behavior has also been studied for marginal detections by Bassett &

Afshordi (2010). To avoid this bias, we recommend that the PDF be estimated, and not assumed Gaussian.

Finally, there is not one 'ideal' statistical method. Different methods have both advantages and disadvantages and a combination of different methods can prove complementary.

4.2. The Saclay method

We present here a new ISW detection method which uses the fields as input (and not the spectra). As we have seen, several statistical tests are interesting in the sense that they do not address the same questions. Therefore we believe that a solid ISW detection method should test:

1. *The correlation detection*: this test is independent of the cosmology.
2. *The amplitude estimation*: this will seek for a specific signal.
3. *The model comparison*: this allows us to check whether the model with ISW is preferred to the model without ISW.

Using the fields instead of the spectra means we must deal with the problem of missing data, which we solve with a sparse inpainting technique (see Appendix A). Such a method has already been applied with success for CMB lensing estimations (Perotto et al. 2010).

The last important issue is how we estimate the final detection level. As explained before, the z-score asymptotically follows a Gaussian distribution and so a bootstrap or Monte-Carlo method is required to derive the correct p -value from the *true* test distribution.

4.2.1. Methodology

Our optimal strategy (summarised by Figure 2) for ISW detection is the following:

1. Apply sparse inpainting to both the galaxy and CMB maps which may have different masks, essentially reconstructing missing data around the Galactic plane and bulge.
2. Test for simple correlation, using a double bootstrap (one to estimate the variance and another to estimate the confidence) and using the fields as input. This returns a model-independent detection level.
3. Reconstruct the ISW signal using expected cosmology and the inpainted galaxy density map.
4. Estimate the signal amplitude using the fields as input, and apply a bootstrap (or MC) to the estimator. This validates both the signal and the model.
5. Apply the model comparison test using the fields as input (see Section 4.2.2), essentially testing whether the data prefers a fiducial (e.g., Λ CDM) model over the null hypothesis, and measure relative fit of models.

As we have chosen to work with the fields, the very first step is to deal with the missing data. This is an ill-posed problem, which can be solved using sparse inpainting (see Appendix A for more details). This approach reconstructs the entirety of the field, including along the Galactic plane and bulge. We show in Section 5 that the use of sparse inpainting does not introduce a bias in the detection of the ISW effect.

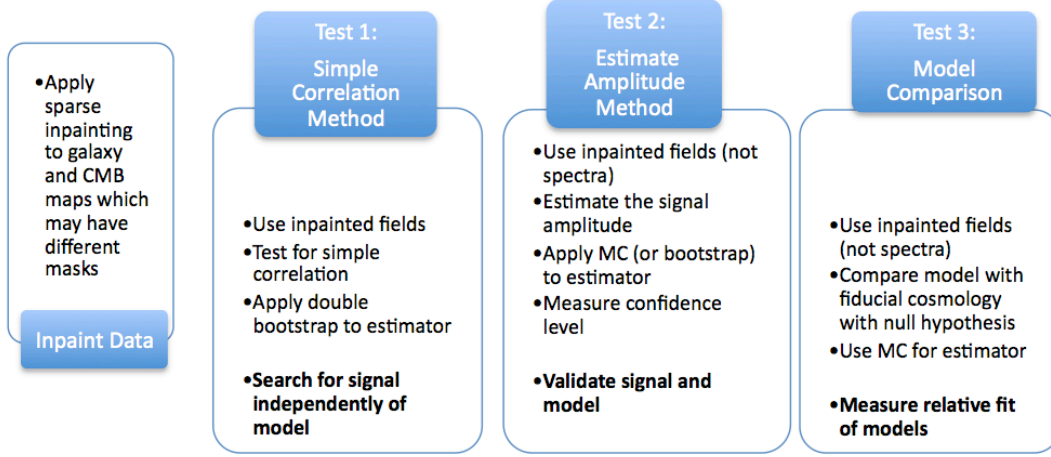


Fig. 2. Description of the steps involved in our method for detecting the ISW effect. Tests 1 to 3 are complementary and ask different statistical questions.

We then perform a correlation detection on the reconstructed fields data using a double bootstrap (see Appendix B.1). Experiments show that the bootstrap tends to over-estimate the confidence interval, especially when the p -values are small. This is why the obtained detection must be used as an indicator when near a significant value (for high p -values, the bootstrap remains accurate).

The second test evaluates the signal amplitude, which validates both the presence of a signal and the chosen model. Bootstrapping test can also be used here as it has no assumption on the underlying cosmology. However, since the accuracy of bootstrap depends on the quantity of observed elements, it may become inaccurate for low p -value, i.e. when there is detection. In such case, Monte-Carlo (MC) will provide more accurate p -values and, for example, with 10^6 MC simulations, we have an accuracy of about 1/1000.

This second test compares the ISW signal with a fiducial model, but does not consider the possibility that the measured signal could in fact be consistent with the ‘null hypothesis’. So even with a significant signal, a third test is necessary. This more pertinent question is addressed by using the ‘Model Comparison’ method (defined in Section 4.2.2), for the first time using the fields approach.

In conclusion, our method consists of a series of complementary tests which together answer several questions. The first test seeks the presence of a correlation between two fields, without any referring cosmology. The second model-dependent test searches a given signal and tests its nullity. The third test asks whether the data prefers a fiducial ISW signal over the null hypothesis.

4.2.2. ‘Field’ model comparison

We define here the model comparison technique using the fields approach, which has until now not been used in the literature. Using a generalised likelihood ratio approach, the quantity to measure is:

$$K_6 = \delta_{\text{OBS}}^* C_{TT}^{-1} \delta_{\text{OBS}} - (\delta_{\text{ISW}} - \delta_{\text{OBS}})^* C_{TT}^{-1} (\delta_{\text{ISW}} - \delta_{\text{OBS}}). \quad (24)$$

Theoretically K_6 converges asymptotically to a χ^2 variable with a certain number of *d.o.f.*s. As we only have one observation we cannot assume (asymptotic) convergence. We can however use a Monte Carlo approach in order to estimate the p -value of the test under the H_0 hypothesis.

The p -value is defined as the probability that under H_0 the test value can be over a given K_6 , i.e. $P(t > K_6) = \int_{K_6}^{\infty} p(x) dx$, where p is the probability distribution of the test under the H_0 hypothesis. By simulating primordial CMB for a fiducial cosmology, these values can be easily computed. Then the p -value gives us a confidence on rejecting the H_0 hypothesis.

Notice that the same procedure for the p -value estimation can be applied on K_4 (Equation 22), even on K_1 (Equation 15) and K_3 (Equation 17) for the power spectra methods. We will further refer to this p -value estimation as the Monte-Carlo estimation, as we theoretically know the distribution under the null hypothesis, i.e. the primordial CMB is supposed to come from a Gaussian random process.

5. Validation of the Saclay Method

In order to validate the Saclay method, we estimate the detection level expected **using WMAP 7 data for the CMB and 2MASS and Euclid data for the galaxy data** (see section 6 for a description of WMAP and 2MASS data sets). We quantify the effect of the inpainting process on CMB maps with and without an ISW signal. We do this by simulating 2MASS-like **and Euclid-like Gaussian and lognormal galaxy distributions** and WMAP7-like Gaussian CMB maps (using cosmological parameters from Table 6) both with and without an ISW signal. We then apply our method to attempt a detection of the ISW signal. We do this both on full-sky maps as well as on masked data where we have reconstructed data behind the mask using the sparse inpainting technique (the masks we use are as described in 6.1 and 6.2).

For each simulation, we run the 3-step Saclay method. Except for the cross-correlation method where we use 100

iterations for the 2MASS-like and 1000 for Euclid-like simulations for the p -value estimation and 201 for the variance estimation (nested bootstrap), every other Monte-Carlo process was performed using 10 000 iterations. All tests were performed inside the spherical harmonics domain with $\ell \in [2, 100]$ for 2MASS and $\ell = [2, 350]$ for a Euclid-like survey.

For the Euclid-like survey, we consider a galaxy distribution as defined in Amara & Réfrégier (2007), with mean redshift $z_m = 0.8$ and slopes $\alpha = 2, \beta = 1.5$. We reconstruct the ISW effect created by the projected galaxy distribution of the Euclid survey, by considering only one large redshift bin. In the future, it could be possible to refine such a reconstructed map by considering tomographic bins, or using information from the spectroscopic survey. As sky coverage maps are not yet available for Euclid, we consider the same mask as for 2MASS and inpaint regions with missing data following Section 6.3. We choose to do this, rather than simply assume a value for the fraction of sky covered (f_{sky}), so as to consider more realistic problems relating to the shape of the mask, and to test our inpainting method.

5.1. Expected level of detection:

The expected detection levels (in units of σ) are reported in Table 3 (2MASS) and Table 4 (Euclid). Methods 1 – 3 correspond to the 3-step method described in Figure 2, where (b) and (MC) denote bootstrap and Monte Carlo evaluations of the variance and the p -value of the test. The p -values are converted as a σ value using the following formula:

$$s = \sqrt{2} \operatorname{erf}^{-1}(1 - p), \quad (25)$$

where p is the p -value, s the corresponding σ -score and erf^{-1} the inverse error function.

The 2MASS simulations (Table 3) show that we expect the same level of significance for an ISW detection, whether the mass tracer follows a Gaussian or a lognormal distribution. In either case the significance is low, around $1\sigma \pm 1\sigma$. This means that for 2MASS-like survey, we have a signal to noise ratio (S/N) around 1σ . We see no major difference between the expected detection levels of M2 and M3. The only difference is for M1 (but there is still agreement with M2 and M3 within 1σ error bars) - this may be due to the fact that the bootstrap technique is more efficient for Gaussian assumptions.

We also apply our method to CMB simulations with no ISW signal present (2 left columns of Table 3), and find a lower detection significance than when an ISW signal is present. This is true even when inpainting is used to recover missing data, showing that the inpainting method does not introduce spurious correlations.

In any case, all methods suggest it is difficult to detect the ISW signal with high significance using the 2MASS data as a local tracer of the matter distribution.

In Table 4, we show that an Euclid-like survey, which is optimally designed for an ISW detection (see, Douspis et al. 2008), permits a much

Gaussian density		
Method	Monte-Carlo	Bootstrap
Cross-Correlation (Equation 11)	0.46 ± 0.28	0.44 ± 0.35
Amplitude estimation (Equation 21)	0.44 ± 0.30	0.47 ± 0.33
Lognormal density		
Method	Monte-Carlo	Bootstrap
Cross-Correlation (Equation 11)	0.39 ± 0.27	0.38 ± 0.31
Amplitude estimation (Equation 21)	0.41 ± 0.26	0.40 ± 0.30

Table 5. Expected p -values for inpainted maps of ISW signal for a 2MASS-like local tracer of mass using Monte-Carlo or bootstrap methods for the first two steps of the Saclay method.

higher detection than with a 2MASS survey. As with 2MASS simulations, we notice that M2 and M3 return similar detection levels, which are lower than M1. Inclusion of masked data reduced the significance, but our inpainting method does not introduce spurious correlations as inpainted maps with no ISW do not return a detection. For a Euclid-like survey with incomplete sky coverage, we can expect to show that the data prefers an ISW component over no dark energy (M3) at the 4.7σ level, and detect a cross-correlation signal at the $\sim 7\sigma$ level. We find no significant differences in the detection levels when the simulations are assumed lognormal or Gaussian.

We also investigate the performance of the wild bootstrap method for the confidence estimation. Table 5 shows the p -values estimated using Monte-Carlo procedure and wild bootstrap for the first two methods of the 3-step Saclay method. Notice that the bootstrap results are almost equivalent to Monte-Carlo ones. We find the bootstrap method wasn't always reliable when the p -value becomes small, because the precision of the bootstrap depends on both the number of bootstrap samples (as any MC-like process) and the number of observed elements. This last dependence makes the bootstrap uncertain when the detection is almost certain, that is why we consider the bootstrapped cross-correlation as an indicator that needs refinement when the results are very significant.

5.2. Power of the tests

In order to investigate the different strengths of each method, we also evaluate the rate of true positives vs. the number of false positives (i.e. false detections) - this information is summarised in Figure 3 which shows Receiver Operating Characteristic (ROC) curves. The construction of the ROC curve requires the computation of p -values for several simulated cases (i.e., simulations with and without an ISW signal), which are then sorted by value. For each p -value or threshold, the corresponding false positive and true positive rates are computed.

Generally, a more sensitive method may be more permissive and so will return a higher proportion of false detections. An ideal method will have a ROC curve above the diagonal from (0, 1) to (1, 1). Similarly, a poor detector will produce a curve below the diagonal, which corresponds to

2MASS / Method		Full sky maps (with ISW)	Inpainted maps (with ISW)	Full sky map (no ISW)	Inpainted map (no ISW)
Gaussian matter density field					
1	Simple correlation (b)	1.12 ± 0.98	1.08 ± 0.99	0.78 ± 0.81	0.81 ± 0.86
2	Field amplitude (MC)	0.94 ± 0.67	0.91 ± 0.65	0.68 ± 0.57	0.71 ± 0.58
3	Field model test (MC)	0.94 ± 0.67	0.91 ± 0.64	0.69 ± 0.57	0.72 ± 0.58
Lognormal matter density field					
1	Simple correlation (b)	1.37 ± 1.51	1.30 ± 1.37	0.85 ± 1.05	0.98 ± 1.34
2	Field amplitude (MC)	0.97 ± 0.62	1.00 ± 0.65	0.69 ± 0.54	0.78 ± 0.61
3	Field model test (MC)	0.97 ± 0.62	1.00 ± 0.65	0.69 ± 0.53	0.78 ± 0.60

Table 3. Expected detection level (units of σ) of ISW signal for a 2MASS-like local tracer of mass assuming Gaussian and lognormal distribution for a fiducial cosmology (see Table 6). Methods 1 – 3 represent the 3-step method. Angular scales included in the analysis are $\ell = [2 - 100]$.

Euclid / Method		Full sky maps (with ISW)	Inpainted maps (with ISW)	Full sky map (no ISW)	Inpainted map (no ISW)
Gaussian matter density field					
1	Simple correlation (b)	> 7	6.98 ± 2.30	0.87 ± 0.67	0.99 ± 0.72
2	Field amplitude (MC)	> 5	4.70 ± 2.40	0.78 ± 0.49	0.88 ± 0.54
3	Field model test (MC)	> 5	4.77 ± 2.42	0.78 ± 0.48	0.88 ± 0.54
Lognormal matter density field					
1	Simple correlation (b)	> 7	6.29 ± 2.60	0.84 ± 0.70	0.77 ± 0.72
2	Field amplitude (MC)	> 5	4.43 ± 2.37	0.75 ± 0.54	0.74 ± 0.54
3	Field model test (MC)	> 5	4.43 ± 2.37	0.74 ± 0.53	0.74 ± 0.54

Table 4. Expected detection level (units of σ) of ISW signal for a Euclid-like local tracer of mass assuming Gaussian distribution for a fiducial cosmology (see Table 6). Methods 1 – 3 represent the 3-step method. Angular scales included in the analysis are $\ell = [2 - 300]$.

odds worse than tossing a coin. **The X-axis corresponds to the false positive rate, i.e. the ratio of CMB maps without ISW where ISW signal is detected at the current threshold. The Y-axis corresponds to the true positive rate, i.e. the ratio of CMB maps with ISW where ISW signal is detected at the current. We recall that a point on the ROC curve corresponds to a threshold.**

Figure 3 shows the ROC curves for three methods applied to a 2MASS-like survey (*Left*) and a Euclid-like survey (*emphRight*): fields model test (thin solid green), field amplitude estimation (dot-dashed blue), simple correlation (dashed red). For the 2MASS survey, all the methods are inside the 1σ error bar of the field’s amplitude estimation and so are nearly equivalent, i.e. no method performs better than the others.

For the Euclid-like survey, the statistics return much better values then for 2MASS (i.e., the ROC curves are far from the diagonal). The simple correlation method will return more false positives than the other two methods, which are nearly identical. The ROC curves for each method differ by more than 1σ at some points and so different methods will perform differently.

6. The ISW signal in WMAP7 due to 2MASS galaxies

We apply the new detection method described in Section 4 to WMAP7 data (Jarosik et al. 2010) and the 2MASS galaxy survey which has been extensively used as a tracer of mass for the ISW signal (see Table 1). We describe first the data in Sections 6.1 and 6.2. In Section 6.3 we describe the

inpainting process that we apply to both CMB and galaxy data. In Section 6.4, we present the detection results.

6.1. WMAP

For the cosmic microwave background data, we use several maps from NASA Wilkinson Microwave Anisotropy Probe: the internal linear combination map (ILC) for years 5 and 7 (WMAP5, Komatsu et al. 2009) (WMAP7, Jarosik et al. 2010) and the ILC map by Delabrouille et al. (2009) which was reconstructed using a needlets technique. We avoid regions which are contaminated by Galactic emission by applying the *Kq85* temperature mask - which roughly corresponds to the *Kp2* mask from the third year release (see Figure 4). **We also subtract the kinetic Doppler quadrupole contribution from the data.** WMAP simulations used to produce Tables 3 and 4 use WMAP 7 best fit parameters for a flat Λ CDM universe (see Table 6).

Ω_b	0.0449	Ω_m	0.266	Ω_Λ	0.734
n	0.963	σ_8	0.801	h	0.710
τ	0.088	w_0	-1.00	w_a	0.0

Table 6. Best fit WMAP 7 cosmological parameters used throughout this paper.

6.2. 2MASS Galaxy Survey

The 2 Micron All-Sky Survey (2MASS) is a publicly available full-sky extended source catalogue (XSC) selected in the near-IR (Jarrett et al. 2000). The near-IR selection

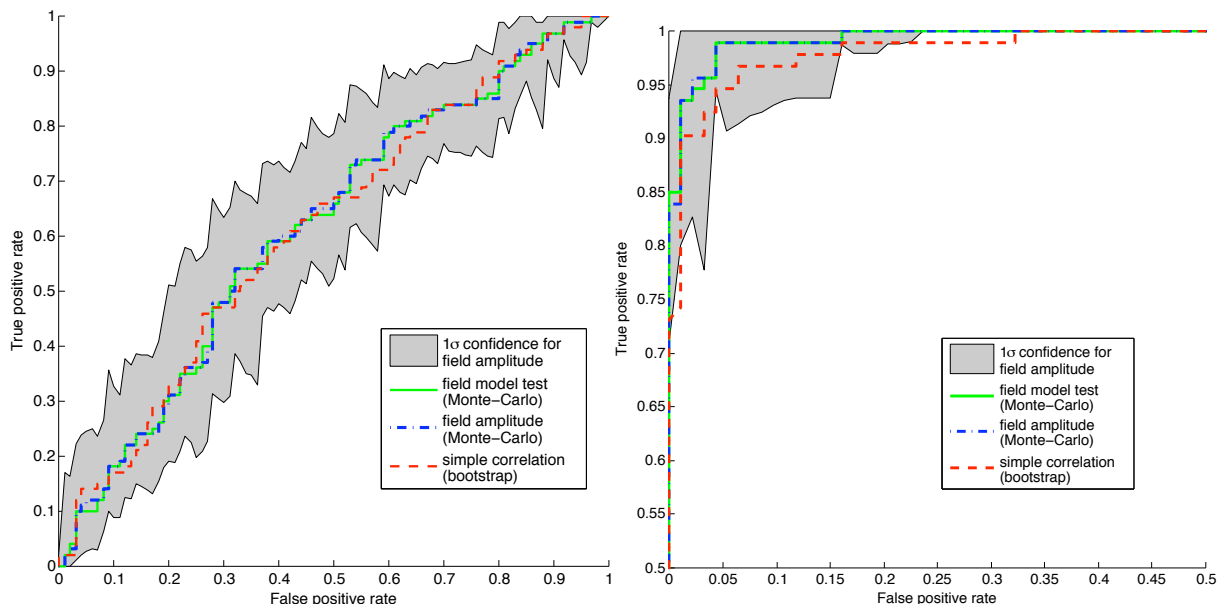


Fig. 3. ROC curves for the 3-step Saclay method for the 2MASS survey (*Left*) and the Euclid survey (*Right*). Note the axes in the right-hand panel are different than in the left-hand panel. The statistical methods correspond to: fields model test (thin solid green, Method 3 in Table 3), field amplitude estimation (dot-dashed blue, Method 2), simple correlation (dashed red, Method 1). For the 2MASS survey, all the methods are inside the 1σ error bar of the field’s amplitude estimation and so are nearly equivalent. For the Euclid-like survey, the statistics return much better values than for 2MASS (i.e., the ROC curves are far from the diagonal). The simple correlation method will return more false positives than the other two methods, which are nearly identical. The ROC curves of the simple correlation test differs sometimes by more than 1σ at some points from the other two methods and so is expected to perform differently.

Method	Inpainted maps (with ISW)
Simple cross-correlation (Eq. 13)	$1.30\sigma \pm 1.37$
Amplitude estimation (Eq. 22)	$1.00\sigma \pm 0.65$
Model selection (Eq. 23)	$1.00\sigma \pm 0.65$

Table 7. Expected detections for 2MASS like survey (from log-normal results in Table 3).

means galaxies are surveyed deep into the Galactic plane, meaning 2MASS has a very large sky coverage, ideal for detecting the ISW signal.

Following (Afshordi et al. 2004), we create a mask to exclude regions of sky where XSC is unreliable using the IR reddening maps of Schlegel et al. (1998). Using $A_k = 0.367 \times E(B - V)$, Afshordi et al. (2004) find a limit $A_K < 0.05$ for which 2MASS is seen to 98% complete for $K_{20} < 13.85$, where K_{20} is the K_s -band isophotal magnitude. Masking areas with $A_K > 0.05$ leaves 69% of the sky and approximately 828 000 galaxies for the analysis (see Figure 4).

We use the redshift distribution computed by Afshordi et al. (2004) (and also used in Rassat et al. 2007), and in order to maximise the signal, we consider one overall bin for magnitudes $12 < K < 14$. The redshift distribution for 2MASS is that shown in Figure 1 of Rassat et al. (2007) (solid black line) and peaks at $z \sim 0.073$. The authors also showed that the small angle approximation could be used for calculations relating to 2MASS, so Equations 10 and 4

can be replaced by their simpler small angle form:

$$C_{gT}(\ell) = \frac{-3bH_0^2\Omega_{m,0}}{c^3(\ell + 1/2)^2} \int dr \Theta D^2 H[f - 1] P\left(\frac{\ell + 1/2}{r}\right), \quad (26)$$

and

$$C_{gg}(\ell) = b^2 \int dr \frac{\Theta^2}{r^2} D^2 P\left(\frac{\ell + 1/2}{r}\right) \quad (27)$$

We estimate the bias from the 2MASS galaxy power spectrum using the cosmology in Table 6 and find $b = 1.27 \pm 0.03$, which is lower than that found in Rassat et al. (2007).

6.3. Applying Sparse Inpainting to CMB and Galaxy data

As discussed in Section 4, regions of missing data in galaxy and CMB maps constitute an ill-posed problem when using a ‘field’ based method. We propose to use sparse inpainting (see Starck et al. 2010, and appendix A) to reconstruct the field in the regions of missing data.

We apply this method to both the WMAP7 and 2MASS maps, and the reconstructed maps are shown in Figure 4. All maps are pixelised using the HEALPix software (Górski et al. 2002; Gorski et al. 2005) with resolution corresponding to NSIDE=512. The top two figures show 2MASS data (left) and ILC map (right) with the masks in grey. The two figures in the middle show the reconstructed density fields for 2MASS data (left) and the ILC map (right). The bottom two figures show the reconstruction of the ISW field using the inpainted 2MASS density map and Equation 20 (for clarity, the first two multipoles ($\ell = 0, 1$) are not present in this map).

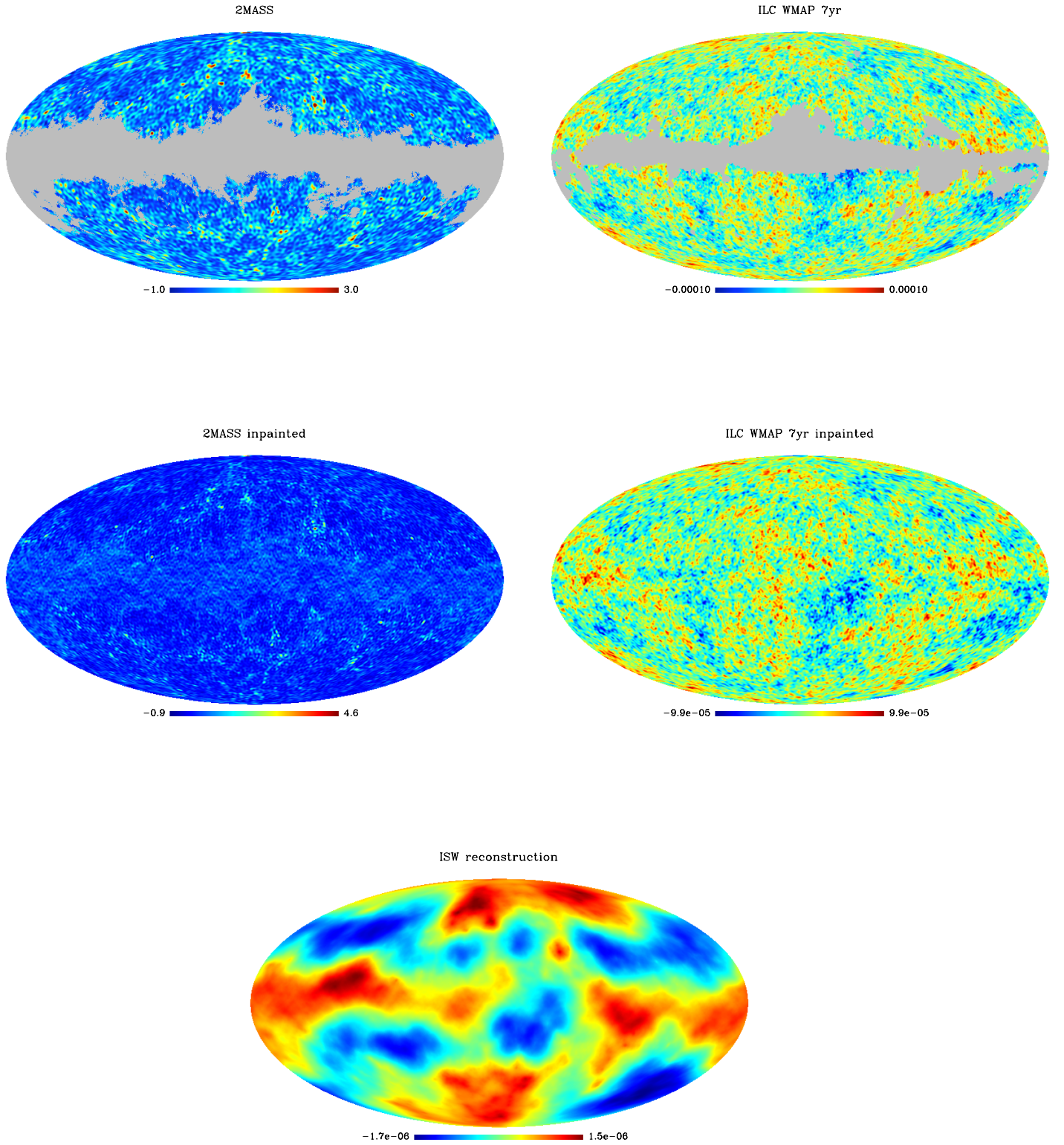


Fig. 4. Top: 2MASS map with mask (left) and WMAP 7 ILC map with mask (right). Middle: Reconstructed 2MASS (left) and WMAP 7 ILC (right) maps using our inpainting method. Bottom: reconstructed ISW temperature field due to 2MASS galaxies, calculated using Equation 20. For better visualisation of the maps as an input of the Saclay method, we consider only the information inside $\ell \in [2, 200]$.

6.4. ISW Detection using 2MASS and WMAP7 Data

We use the Saclay method (section 4) on WMAP7 and 2MASS data with 10^6 iterations for the Monte-Carlo iterations, 1000 iterations for both bootstrap and nested bootstrap of the correlation detection, and search for the ISW signal. **In order to test the effect of including smaller (and possibly non-linear) scales, we perform the analysis for three different ℓ -ranges: $\ell \in [\ell_{\min}, 50], [\ell_{\min}, 100], [\ell_{\min}, 200]$, where $\ell_{\min} = 2$ or 3. We notice that inclusion or not of the quadrupole ($\ell = 2$) can affect the significance level slightly, which is why we chose two values for ℓ_{\min} .** We report our measured detection levels in Table 8. We recall the expected detections for 2MASS in Table 7, which can also be found in the more complete set of results presented in Table 3.

The results in Table 8 are compatible with the results in Table 3 within the errors bars. Rassat et al. (2007) used a spectra model-comparison method over $\ell = [3 - 30]$, and found that the model with dark energy was marginally preferred over the null hypothesis. Over a range $\ell = [2/3 - 50]$, using a fields model-comparison method, we find that a model with dark energy is preferred over the null hypothesis at the $1.1 - 1.8\sigma$ level, depending on the map. More generally, our results are compatible with the earliest ISW measurement using 2MASS data (Afshordi et al. 2004; Rassat et al. 2007) and lie in the $1.1 - 2.0\sigma$ range depending on the data and statistical test used.

The simple correlation test tends to report marginally higher detection levels than the field amplitude and model comparison tests, and the model comparison test similar values as the field amplitude test, **which is compatible with the predictions from the ROC curves in Figure 3.**

7. Discussion

In this paper we have extensively reviewed the numerous methods in the literature which are used to detect and measure the presence of an ISW signal using maps for the CMB and local tracers of mass. We noticed that the variety of methods used can lead to different and conflicting conclusions. We also noted two broad classes of methods: one which uses the cross-correlation spectrum as the measure and the other which uses the reconstructed ISW temperature field.

We identified the advantages and disadvantages of all methods used in the literature and concluded that:

1. Using the fields (instead of spectra) as input required *only* the primordial CMB to be Gaussian. This requires reconstruction of the ISW field, which is difficult with missing data.
2. The ill-posed problem of missing data can be solved using sparse inpainting, **a method which does not introduce spurious correlations between maps.**
3. Assuming the estimator was Gaussian led to underestimation of the signal estimation.
4. A series of statistical tests could provide complementary information.

This led us to construct a new and complete method for detecting and measuring the ISW effect. The method is summarised as follows:

1. Apply sparse inpainting to both the galaxy and CMB maps which may have different masks, essentially reconstructing missing data around the Galactic plane and bulge.
2. Test for simple correlation, using a double bootstrap (one to estimate the variance and another to estimate the confidence) and using the fields as input. This returns a model-independent detection level.
3. Reconstruct the ISW signal using expected cosmology and the inpainted galaxy density map;
4. Estimate the signal amplitude using the fields as input, and apply a bootstrap (or MC) to the estimator. This validates both the signal and the model.
5. Apply the fields model comparison test, essentially testing whether the data prefers a given model over the null hypothesis, and measure relative fit of models.

The method we present in this paper makes only one assumption: that the primordial CMB temperature field behaves like a Gaussian random field. The method is general in that it ‘allows’ the galaxy field to behave as a lognormal field, but does not automatically assume that the galaxy field is lognormal.

We first applied our method to 2MASS and Euclid simulations. We find that it is difficult to detect the ISW significantly using 2MASS simulations, and find no difference between assuming the underlying galaxy field is Gaussian or lognormal, and only mild differences depending on the statistical test used. With a Euclid-like survey, we expect high detection levels, even with incomplete sky coverage - we expect $\sim 7\sigma$ detection level using the simple correlation method, and $\sim 4.7\sigma$ detection level using the fields amplitude or method comparison techniques. These detection levels are the same whether the Euclid galaxy field follows a Gaussian or lognormal distribution. Our results also show that the inpainting method does not introduce spurious correlations between maps.

We applied this method to WMAP7 and 2MASS data, and found that our results were comparable with early detections of the ISW signal using 2MASS data (Afshordi et al. 2004; Rassat et al. 2007) and lied roughly in the $1.1 - 2.0\sigma$ range. **These results are also compatible with the simulations we ran for the 2MASS survey.**

The last test we performed, the model comparison test, asks the much more pertinent question of whether the data prefers a Λ CDM model to the null hypothesis (i.e. no curvature and no dark energy). Using this test, we find a $1.1 - 1.8\sigma$ detection for ranges $\ell = [2/3 - 50]$ and $1.2 - 1.9\sigma$ for ranges $\ell = [2/3 - 100/200]$, which is sometimes higher than what was previously reported in Rassat et al. (2007) using a spectra models comparison test, without sparse inpainting or bootstrapping. **A by-product of this measurement is the reconstruction of the temperature ISW field due to 2MASS galaxies, reconstructed with full sky coverage.**

By applying our method on different estimation of the CMB map, we have highlighted the effect of the component separation on the ISW detection. Table 8 shows score between $1.1 - 2.0\sigma$ on that should be almost the same data, and with a previous test not reported here, we were able to detect ISW at

Galaxy Survey	2MASS		
	WMAP ILC 7yr	WMAP ILC 5yr	Needlets ILC 5yr
Scales interval	$\ell \in [2, 50]$		
Simple correlation detection	1.3σ	1.3σ	1.2σ
Field amplitude	1.3σ	1.3σ	1.1σ
Model comparison	1.3σ	1.3σ	1.1σ
Scales interval	$\ell \in [2, 100]$		
Simple correlation detection	1.5σ	1.4σ	1.3σ
Field amplitude	1.4σ	1.3σ	1.2σ
Model comparison	1.4σ	1.3σ	1.2σ
Scales interval	$\ell \in [2, 200]$		
Simple correlation detection	1.4σ	1.4σ	1.3σ
Field amplitude	1.4σ	1.3σ	1.2σ
Model comparison	1.4σ	1.3σ	1.2σ
Scales interval	$\ell \in [3, 50]$		
Simple correlation detection	2.0σ	1.9σ	1.5σ
Field amplitude	1.8σ	1.7σ	1.4σ
Model comparison	1.8σ	1.7σ	1.4σ
Scales interval	$\ell \in [3, 100]$		
Simple correlation detection	2.0σ	2.0σ	1.5σ
Field amplitude	1.9σ	1.7σ	1.5σ
Model comparison	1.9σ	1.7σ	1.5σ
Scales interval	$\ell \in [3, 200]$		
Simple correlation detection	1.9σ	1.9σ	1.6σ
Field amplitude	1.8σ	1.7σ	1.5σ
Model comparison	1.8σ	1.7σ	1.5σ

Table 8. Detection levels obtained with different CMB maps and different scales intervals (considering with and without quadrupole) using the Saclay method described in 4.

2.7σ using a another map (not presented in this paper). The influence of the component separation method on the quality of the estimation needs to be more deeply understood in future works.

Acknowledgements. This work has been supported by the European Research Council grant SparseAstro (ERC-228261). We thank Ofer Lahav, John Peacock and the anonymous referee for useful discussions. We also thanks Jacques Delabrouille for providing us the needlets ILC 5yr map from the WMAP 5yr data. We used Healpix software (Górski et al. 2002; Gorski et al. 2005), ISAP¹ software, the 2MASS catalogue², the WMAP data³ and the Galaxy extinction maps of Schlegel et al. (1998).

Appendix A: Sparse Inpainting

The goal of inpainting is to restore missing or damaged regions of an image, in such a way that the restored map has the same statistical properties as the underlying unmasked map (Elad et al. 2005). Sparse Inpainting has been proposed for filling the gaps in CMB maps (Abrial et al. 2007, 2008) and for weak lensing mass map reconstruction (Pires et al. 2009, 2010). In Perotto et al. (2010), it has been shown that the sparse inpainting method does not destroy the CMB weak lensing signal, and is therefore an elegant way to handle the mask problem.

The inpainting problem can be defined as follows. Let X be the ideal complete image, Y the observed incomplete image (images can be fields on the sphere) and L the binary mask (i.e. $L[k, l] = 1$ if we have information at pixel

(k, l) , $L[k, l] = 0$ otherwise). In short, we have: $Y = LX$. Inpainting consists in recovering X knowing Y and L . The masking effect can be thought of as a loss of sparsity in the spherical harmonic domain since the information required to define the map has been spread across the spherical harmonic basis.

Sparsity means that most of the information is concentrated in a few coefficients, which when sorted from the largest to the smallest, follow an exponential decay. More details can be found in (Starck et al. 2010). In this paper, the chosen ‘dictionary’ is the spherical harmonic domain.

Denoting the spherical harmonic basis as Φ (so Φ^T is the spherical harmonic transform, i.e. the projector onto the spherical harmonic space), $\|z\|_0$ the l_0 pseudo-norm, i.e. the number of non-zero entries in z and $\|z\|$ the classical l_2 norm (i.e. $\|z\|^2 = \sum_k (z_k)^2$), we thus want to minimise:

$$\min_X \|\Phi^T X\|_0 \quad \text{subject to} \quad \|Y - LX\|_{\ell_2} \leq \sigma, \quad (\text{A.1})$$

where σ stands for the noise standard deviation in the noisy case. It has also been shown that if X is sparse enough, the l_0 pseudo-norm can also be replaced by the convex l_1 norm (i.e. $\|z\|_1 = \sum_k |z_k|$) (Donoho & Huo 2001). The solution of such an optimisation task can be obtained through an iterative thresholding algorithm called MCA (Elad et al. 2005; Fadili et al. 2009; Starck et al. 2010) :

$$X^{n+1} = \Delta_{\Phi, \lambda_n}(X^n + Y - LX^n) \quad (\text{A.2})$$

where the nonlinear operator $\Delta_{\Phi, \lambda}(Z)$ consists in:

1. decomposing the signal Z on the dictionary Φ to derive the coefficients $\alpha = \Phi^T Z$ (i.e. α is the vector containing the $a_{\ell m}$ coefficients).

¹ <http://jstarck.free.fr/isap.html>

² <http://www.ipac.caltech.edu/2mass/>

³ <http://map.gsfc.nasa.gov>

2. threshold the coefficients: $\tilde{\alpha} = \rho(\alpha, \lambda)$, where the thresholding operator ρ can either be a hard thresholding (i.e. $\rho(\alpha_i, \lambda) = \alpha_i$ if $|\alpha_i| > \lambda$ and 0 otherwise) or a soft thresholding (i.e. $\rho(\alpha_i, \lambda) = \text{sign}(\alpha_i)\max(0, |\alpha_i| - \lambda)$). The hard thresholding corresponds to the l_0 optimisation problem while the soft-threshold solves that for l_1 .
3. reconstruct \tilde{Z} from the thresholds coefficients $\tilde{\alpha}$.

The threshold parameter λ_n decreases with the iteration number and it plays a role similar to the cooling parameter of the simulated annealing techniques, i.e. it allows the solution to escape from local minima.

It has been shown in Abrial et al. (2008) that this inpainting technique leads to accurate CMB recovery results. More details relative to this optimisation problem can be found in Combettes & Wajs (2005); Starck et al. (2010) and theoretical justifications for CMB sparse inpainting in Rauhut & Ward (2010). The software is available in the Multi-Resolution on the Sphere (MRS) package ⁴.

Appendix B: Bootstrapping detection tests

Most detection methods rely on a z-score whose confidence is computed assuming a Gaussian distribution (see section 3.2). This assumption is valid within the central limit theorem and so is not always true. Using an estimator of the true distribution will give more reliable confidence levels. To estimate the true distribution, two approaches are generally used: Monte-Carlo simulations when a model and its parameters are available or non-parametric bootstraps methods.

In order to build a fully non-parametric and cosmology independent correlation detection test, we propose to use the wild bootstrap in order to estimate both estimator variance and confidence interval. The remaining question will be the reliability of such test when bootstraps are sometimes known to underestimate confidence intervals. The answer to the question depends mainly on the choice of domain (e.g. pixels, spherical harmonics) and the method chosen for estimating.

Working with correlated data is difficult and estimated values are generally different to the true ones. We chose to work in the spherical harmonic domain, which produces heteroskedastic (i.e., the power spectrum is scale dependent) and uncorrelated data. For heteroskedastic and uncorrelated data, bootstraps methods like wild bootstrap can be used. Flachaire (2005); Davidson & Flachaire (2008) showed the good behaviour of wild bootstrap in many situations. We propose here a short introduction to this method, which we use to estimate p -values and estimator variances.

B.1. Wild Bootstrap and Regression Model

Bootstrap methods were introduced by Efron (1979) as a generalisation of the jackknife and its concept has been extended to many situations. For the regression case with heteroskedastic data, wild bootstrap was developed by Liu (1988), who established the ability of wild bootstrap to provide refinements for the linear regression model with heteroskedastic disturbances. Mammen (1993) showed that the wild bootstrap was asymptotically justified, in the sense that the asymptotic distribution of various statistics is the

same as the asymptotic distribution of their wild bootstrap counterparts.

If you look at the correlation estimator (Equation 11) and the amplitude estimation (Equation 21), they both suppose an underlying linear model:

$$\delta_{\ell m}^O = \rho \delta_{\ell m}^C + \sigma_{\ell m}, \quad (\text{B.1})$$

where δ^O is the observed field which contains the signal of interest (e.g. CMB), δ^C a tracer of the signal of interest (e.g. matter density, ISW reconstruction), ρ is either the correlation coefficient and the amplitude of the signal (depending of the estimator) and σ is the noise which is dependent of the position.

The originality of the wild bootstrap relies in the fact that its data generating process (DGP) creates new samples without any prior on the distribution of the noise. This process uses the residual, i.e. $u_{\ell m} = \delta_{\ell m}^O - \hat{\rho} \delta_{\ell m}^C$, where $\hat{\rho}$ is the parameter estimate. Then, new samples are created using the following formula,

$$\delta_{\ell m}^{O*} = \hat{\rho} \delta_{\ell m}^C + u_{\ell m}^*, \quad (\text{B.2})$$

$$u_{\ell m}^* = \frac{u_{\ell m} \varepsilon}{1 - h_{\ell m}}, \quad h_{\ell m} = \frac{|\delta_{\ell m}^C|^2}{\sum_{\ell, m} |\delta_{\ell m}^C|^2} \quad (\text{B.3})$$

where * indicate the bootstrapped version of the sample and ε is a random variable following a given distribution. The performance of the wild bootstrap mainly depends on the distribution of ε , which must verify some conditions like having a null mean and a unit variance.

For example Liu (1988) proposed to use Rademacher variables for the distribution,

$$\varepsilon = \begin{cases} 1 & \text{with probability } 0.5, \\ -1 & \text{with probability } 0.5. \end{cases} \quad (\text{B.4})$$

Davidson & Flachaire (2008) showed that exact inference (up to a theoretical accuracy) is possible if $\delta_{\ell m}^C$ is independent from all the disturbance $u_{\ell m}$ and if the test distribution is symmetric about 0. Moreover, the rate of convergence of the error in the rejection probabilities (ERP) is at most $n^{-3/2}$ with symmetric errors and $n^{-1/2}$ with asymmetric errors (n is the size of the observed sample). The ERP is the difference between the actual rejection frequency under the null hypothesis and the level of the test (e.g. about 0.045 for a 2σ detection). In other words, this is the precision error on the inferred p -value. Notice that it mainly depends on the sample size.

B.2. Parametric bootstrap

The wild bootstrap is very useful when little is known about the noise. However, sometimes the noise distribution is well known and its parameters can be estimated from the observations. In this case, it is possible to use this distribution in order to generate new samples. For example, if we model the CMB primordial distribution by a multivariate normal law, the $u_{\ell m}^*$ can be generated as, $u_{\ell m}^* \sim \mathcal{N}(0, C_{\text{CMB}})$ where C_{CMB} is the power spectra of the primordial CMB. Notice, that using the theoretical CMB power spectra leads to a Monte-Carlo methods which is then dependent on the chosen cosmology.

⁴ <http://jstarck.free.fr/mrs.html>

B.3. Estimating the estimator variance

When the variance of an estimator is unknown, a bootstrap can be used to estimate it with the following scheme:

1. Estimate the parameter vector $\hat{\rho}$ from the observed data;
2. Create B bootstrap samples, using Equation B.2 or another method (see section B.2);
3. For each bootstrap sample, estimate the parameter vector ρ_i^* , $i \in [1, B]$;
4. Compute the estimation of the variance of the estimator,

$$\text{Var}(\hat{\rho}) = \frac{1}{B-1} \sum_{i=1}^B (\rho_i^* - \bar{\rho}^*)^2, \quad \bar{\rho}^* = \frac{1}{B} \sum_{i=1}^B \rho_i^*. \quad (\text{B.5})$$

B.4. Estimating the p -value of a test

The estimation of the confidence interval of a hypothesis test requires the generation of samples under the null hypothesis. Inside the wild bootstrap DGP, that leads to set $\hat{\rho} = 0$ when using Equation B.2 for new samples (it also remain true with a parametric bootstrap). Using this property, the scheme for computing the p -value is the following,

1. Estimate the parameter vector and its variance (e.g. using the previous scheme in B.3), $\hat{\rho}$ and $\text{Var}(\hat{\rho})$;
2. Compute the z-score,

$$\hat{\tau} = \hat{\rho} / \sqrt{\text{Var}(\hat{\rho})}; \quad (\text{B.6})$$

3. Create B bootstrap samples which follow the null hypothesis using B.2 (or section B.2) and $\hat{\rho} = 0$;
4. For each bootstrap sample, estimate the parameter, ρ_i^* , and its variance $\text{Var}(\rho_i^*)$, $i \in [1, B]$;
5. For each bootstrap sample, compute the correspondent z-score,

$$\tau_i^* = \rho_i^* / \sqrt{\text{Var}(\rho_i^*)}; \quad (\text{B.7})$$

6. Compute the bootstrapped p -value,

$$p^* = \frac{1}{B} \sum_i \mathbf{I}(\tau_i^* > \hat{\tau}), \quad (\text{B.8})$$

where \mathbf{I} denotes the indicator function, which is equal to 1 when its argument is true and 0 otherwise.

Equation B.8 can be a little too rough and may lead to inaccurate results when the answer is just between two bootstrap samples. One way to reduce this effect is by smoothing the empirical distribution (Racine & MacKinnon 2007), for example with a Gaussian kernel,

$$p^* = 1 - \frac{1}{B} \sum_{i=0}^{B-1} \varphi\left(\frac{\hat{\tau} - \tau_i^*}{h}\right), \quad (\text{B.9})$$

where φ is the Gaussian cumulative distribution function and h the bandwidth of the kernel (chosen depending on the required precision, see Racine & MacKinnon 2007).

B.5. Reliability of bootstrap

One can be perplexed by the efficiency and the reliability of bootstrap method. Many studies have proved that these methods outperform asymptotic tests (i.e. using the asymptotic distribution) (Davidson & MacKinnon 2006) and sometimes perform almost as well as exact tests (i.e. with the true distribution) (we refer the reader to Efron 1987; DiCiccio & Efron 1996; Hall 1995, for more information). The efficiency is mostly dependent on three key elements: 1) the Data Generating Process (DGP) under both hypotheses (often generating under the alternative hypothesis is non trivial), 2) the number of observations and 3) the number of bootstrapped samples. The variance on p is given by $p(1-100)/B$, where B is the number of bootstraps. While the number of bootstrapped sample can be easily tackled, the DGP on the chosen method and the number of observation depends on many parameters (e.g. configuration, instruments, acquisition time). Even with the most efficient DGP, the precision of the bootstrap methods is limited by the number of observations, often in $\mathcal{O}(1/\sqrt{n})$, sometimes $\mathcal{O}(1/n)$. In other words, bootstrap methods are mostly inaccurate when the observed value is at the tail of the involved distribution (i.e. and so rare event). For example, a p -value estimated using 100 observations will be ± 0.1 and be inefficient for characterizing high confidence, but can still be useful for rejecting.

References

- Abrial, P., Moudden, Y., Starck, J., et al. 2007, *J. Fourier Anal. Appl.*, 13, 729–748
- Abrial, P., Moudden, Y., Starck, J.-L., et al. 2008, *Statistical Methodology*, 5, 289
- Adelman-McCarthy, J. K., Agüeros, M. A., Allam, S. S., et al. 2008, *ApJS*, 175, 297
- Afshordi, N., Loh, Y.-S., & Strauss, M. A. 2004, *Phys. Rev.*, D69, 083524
- Agüeros, M. A., Anderson, S. F., Margon, B., et al. 2006, *AJ*, 131, 1740
- Albrecht, A., Bernstein, G., Cahn, R., et al. 2006, *ArXiv Astrophysics e-prints*
- Amara, A. & Réfrégier, A. 2007, *MNRAS*, 381, 1018
- Anderson, S. F., Fan, X., Richards, G. T., et al. 2001, *AJ*, 122, 503
- Bassett, B. A. & Afshordi, N. 2010, *ArXiv e-prints*
- Bennett, C. L., Smoot, G. F., & Kogut, A. 1990, in *Bulletin of the American Astronomical Society*, Vol. 22, Bulletin of the American Astronomical Society, 1336+
- Bernardeau, F., Colombi, S., Gaztañaga, E., & Scoccimarro, R. 2002, *Phys. Rep.*, 367, 1
- Boldt, E. 1987, *Phys. Rep.*, 146, 215
- Boughn, S. & Crittenden, R. 2004, *Nature*, 427, 45
- Boughn, S. P. & Crittenden, R. G. 2002, *Physical Review Letters*, 88, 021302
- Boughn, S. P. & Crittenden, R. G. 2005, *New Astron. Rev.*, 49, 75
- Cabré, A., Fosalba, P., Gaztañaga, E., & Manera, M. 2007, *MNRAS*, 381, 1347
- Cabré, A., Gaztañaga, E., Manera, M., Fosalba, P., & Castander, F. 2006, *MNRAS*, 372, L23
- Cai, Y., Cole, S., Jenkins, A., & Frenk, C. S. 2010, *MNRAS*, 407, 201
- Carroll, S. M. et al. 2005, *Phys. Rev.*, D71, 063513
- Cole, S., Norberg, P., Baugh, C. M., et al. 2001, *MNRAS*, 326, 255
- Combettes, P. L. & Wajs, V. R. 2005, *SIAM Multi. Model. Simul.*, 4, 1168–1200
- Condon, J. J., Cotton, W. D., Greisen, E. W., et al. 1998, *AJ*, 115, 1693

- Corasaniti, P., Giannantonio, T., & Melchiorri, A. 2005, *Phys. Rev. D*, 71, 123521
- Crittenden, R. G. & Turok, N. 1996, *Phys. Rev. Lett.*, 76, 575
- Davidson, A. & MacKinnon, J. G. 2006, *Journal of Econometrics*, 133, 421
- Davidson, R. & Flachaire, E. 2008, *Journal of Econometrics*, 146, 162
- Delabrouille, J., Cardoso, J., Le Jeune, M., et al. 2009, *A&A*, 493, 835
- DiCiccio, T. J. & Efron, B. 1996, *Statistical Science*, 11, 189
- Dodelson, S. 2003, *Modern cosmology (Modern cosmology / Scott Dodelson. Amsterdam (Netherlands): Academic Press. ISBN 0-12-219141-2, 2003, XIII + 440 p.)*
- Donoho, D. & Huo, X. 2001, *IEEE Trans. Info. Theo.*, 47, 2845–2862
- Doroshkevich, A., Tucker, D. L., Allam, S., & Way, M. J. 2004, *A&A*, 418, 7
- Douspis, M., Castro, P. G., Caprini, C., & Aghanim, N. 2008, *AAP*, 485, 395
- Efron, B. 1979, *Annals of Statistics*, 7, 1
- Efron, B. 1987, *The Jackknife, the Bootstrap, and Other Resampling Plans, CBMS-NSF Regional Conference Series in Applied Mathematics (Society for Industrial Mathematics)*
- Efstathiou, G. 2004, *MNRAS*, 349, 603
- Elad, M. and Starck, J.-L., Querre, P., & Donoho, D. 2005, *Applied and Computational Harmonic Analysis*, 19, 340
- Elad, M., Starck, J.-L., Donoho, D., & Querre, P. 2005, *J. Appl. Comp. Harm. Anal.*, 19, 340–358
- Fadili, M. J., Starck, J.-L., & Murtagh, F. 2009, *The Computer Journal*, 52, 64–79
- Flachaire, E. 2005, *Computational Statistics & Data Analysis*, 49, 361, 2nd CSDA Special Issue on Computational Econometrics
- Fosalba, P. & Gaztañaga, E. 2004, *MNRAS*, 350, L37
- Fosalba, P., Gaztañaga, E., & Castander, F. 2003, *ApJ*, 597, L89
- Francis, C. L. & Peacock, J. A. 2010a, *MNRAS*, 406, 14
- Francis, C. L. & Peacock, J. A. 2010b, *MNRAS*, 406, 2
- Frommert, M., Enßlin, T. A., & Kitaura, F. S. 2008, *MNRAS*, 391, 1315
- Gaztañaga, E., Manera, M., & Multamaki, T. 2006, *MNRAS*, 365, 171
- Giannantonio, T., Crittenden, R. G., Nichol, R. C., et al. 2006, *Phys. Rev. D*, 74, 063520
- Giannantonio, T., Scranton, R., Crittenden, R. G., et al. 2008, *Phys. Rev. D*, 77, 123520
- Górski, K. M., Banday, A. J., Hivon, E., & Wandelt, B. D. 2002, in *Astronomical Society of the Pacific Conference Series*, Vol. 281, *Astronomical Data Analysis Software and Systems XI*, ed. D. A. Bohlender, D. Durand, & T. H. Handley, 107–+
- Gorski, K. M. et al. 2005, *ApJ*, 622, 759
- Granett, B. R., Neyrinck, M. C., & Szapudi, I. 2009, *ApJ*, 701, 414
- Hadamard, J. 1902, *Sur les problèmes aux dérivées partielles et leur signification physique.*
- Hall, P. 1995, *The Bootstrap and Edgeworth Expansion*, Springer Series in Statistics (Springer)
- Hamimeche, S. & Lewis, A. 2008, *Phys. Rev. D*, 77, 103013
- Hernández-Monteagudo, C. 2008, *A&A*, 490, 15
- Hernández-Monteagudo, C. 2009, *ArXiv e-prints*
- Hivon, E., Górski, K. M., Netterfield, C. B., et al. 2002, *ApJ*, 567, 2
- Ho, S., Hirata, C., Padmanabhan, N., Seljak, U., & Bahcall, N. 2008, *Phys. Rev. D*, 78, 043519
- Jarosik, N., Bennett, C. L., Dunkley, J., et al. 2010, *ArXiv e-prints*
- Jarrett, T. H. et al. 2000, *Astron. J.*, 119, 2498
- Kamionkowski, M. & Spergel, D. N. 1994, *ApJ*, 432, 7
- Kayo, I., Taruya, A., & Suto, Y. 2001, *ApJ*, 561, 22
- Kinkhabwala, A. & Kamionkowski, M. 1999, *Physical Review Letters*, 82, 4172
- Komatsu, E., Dunkley, J., Nolta, M. R., et al. 2009, *ApJS*, 180, 330
- Liu, R. Y. 1988, *Annals of Statistics*, 16, 1696
- López-Corredoira, M., Sylos Labini, F., & Betancort-Rijo, J. 2010, *A&A*, 513, A3+
- Maddox, S. J., Efstathiou, G., Sutherland, W. J., & Loveday, J. 1990, *MNRAS*, 242, 43P
- Mammen, E. 1993, *Annals of Statistics*, 21, 255
- McEwen, J. D., Vielva, P., Hobson, M. P., Martínez-González, E., & Lasenby, A. N. 2007, *MNRAS*, 376, 1211
- McEwen, J. D., Wiaux, Y., Hobson, M. P., Vanderghenst, P., & Lasenby, A. N. 2008, *MNRAS*, 384, 1289
- Munshi, D., Valageas, P., Cooray, A., & Heavens, A. 2009, *ArXiv e-prints*
- Nolta, M. R. et al. 2004, *ApJ*, 608, 10
- Padmanabhan, N. et al. 2005, *Phys. Rev.*, D72, 043525
- Peacock, J. A., Schneider, P., Efstathiou, G., et al. 2006, *ESA-ESO Working Group on "Fundamental Cosmology"*, Tech. rep.
- Percival, W. J., Nichol, R. C., Eisenstein, D. J., et al. 2007 a, *ApJ*, 657, 51
- Perotto, L., Bobin, J., Plaszczyński, S., Starck, J., & Lavabre, A. 2010, *A&A*, 519, A4+
- Pietrobon, D., Balbi, A., & Marinucci, D. 2006, *Phys. Rev. D*, 74, 043524
- Pires, S., Starck, J., Amara, A., et al. 2009, *MNRAS*, 395, 1265
- Pires, S., Starck, J., & Refregier, A. 2010, *IEEE Signal Processing Magazine*, 27, 76
- Raccanelli, A., Bonaldi, A., Negrello, M., et al. 2008, *MNRAS*, 386, 2161
- Racine, J. S. & MacKinnon, J. G. 2007, *Computational Statistics & Data Analysis*, 51, 5949
- Rassat, A., Land, K., Lahav, O., & Abdalla, F. B. 2007, *MNRAS*, 377, 1085
- Rauhut, H. & Ward, R. 2010, *ArXiv e-prints*
- Refregier, A., Amara, A., Kitching, T. D., et al. 2010, *ArXiv e-prints*
- Sachs, R. K. & Wolfe, A. M. 1967, *ApJ*, 147, 73
- Sawangwit, U., Shanks, T., Cannon, R. D., et al. 2010, *MNRAS*, 402, 2228
- Schaefer, B. M., Fotios Kalovidouris, A., & Heisenberg, L. 2010, *ArXiv e-prints*
- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*, 500, 525
- Schrabback, T., Hartlap, J., Joachimi, B., et al. 2010, *A&A*, 516, A63+
- Scranton, R. et al. 2003, *ArXiv Astrophysics e-prints*
- Song, Y.-S., Hu, W., & Sawicki, I. 2007, *prd*, 75, 044004
- Spergel, D. N. et al. 2003, *ApJS*, 148, 175
- Starck, J.-L., Murtagh, F., & Fadili, M. 2010, *Sparse Image and Signal Processing (Cambridge University Press)*
- Vielva, P., Martínez-González, E., & Tucci, M. 2006, *MNRAS*, 365, 891
- Wild, V., Peacock, J. A., Lahav, O., et al. 2005, *MNRAS*, 356, 247
- Xia, J., Viel, M., Baccigalupi, C., & Matarrese, S. 2009, *J. Cosmology Astropart. Phys.*, 9, 3

Answer to the second report on the Paper: “An Optimal Approach for Measuring the Integrated Sachs Wolfe Effect”

F.-X. Dupé, A. Rassat, J.-L. Starck and M. J. Fadili

We thank the reviewer for their comments and the very detailed review of our article they have provided.

Before answering the different concerns of the reviewer, we summarise here some major changes to the paper:

- Further analysis of the GMCA map has highlighted aspects of the component separation which we do not fully understand, and so we have decided to exclude the GMCA map from our current simulations and data analysis with 2MASS. We will investigate the GMCA data more thoroughly in a future study. As the GMCA map originally returned the highest detection levels, this means that the original claims of higher detection (2.7σ) in the first version of this paper no longer hold. In order to investigate other data sets too, we have also included the Delabrouille et al 2009 ILC map reconstructed using needlets as a new data set with which to detect the ISW signal. Figure 4 and Table 8 have been updated accordingly.
- We have performed a new set of simulations for a Euclid-like survey, which better illustrates the differences between the statistical tests which form the Saclay method. A new Table 4 has been added.
- We have updated our ISW detection simulations. We now consider both Gaussian and lognormal cases for the 2MASS and Euclid galaxy distributions. We find that this has very little effect on the detection levels, since our methods assume only that the CMB field is Gaussian.

We have also corrected all the mistakes pointed out by the reviewer. We present below our answers to the different concerns of the reviewer.

1 Part I, (first level of concern)

*The theory makes a clear cut prediction in this respect: the part of the ϕ that generates ISW is first order in perturbation theory and, in the absence of non-Gaussianities in the primordial inflaton field (as predicted by most of inflationary models), that same part of ϕ should be strictly Gaussian. On the scales of ISW analyses, non-linear corrections should be small *and* with opposite sign in the*

phi dot (see, for instance, the works of Cui et al. 2008, Smith et al. 2009, based upon the analyses of numerical simulations). Although depending on the amount of galaxies/galaxy number density, for the particular case of 2MASS Poisson shot noise should be negligible on the largest scales: this can be deduced by a simple analysis of C_ℓ 's from 2MASS. Only on the smallest scales the C_ℓ 's should become constant (and dominated by the Poisson component).

For clarity throughout this answer, we call ‘ISW’ the secondary temperature field which is positively correlated with the density field, and ‘Rees-Sciama’, the secondary temperature field which is negatively correlated with the density field.

We agree with the referee that the part of the potential which creates the ISW signal is first order in perturbation theory. However, even in the absence of non-Gaussianities in the primordial inflaton field, the potential ϕ will evolve with redshift, and can become lognormal on quasi-linear scales, where the potential is still decaying: this will produce an ISW signal, which will approach Gaussianity on the largest scales and contain some traces of non-Gaussianity on smaller (but not non-linear) scales (see [5, 4]). Furthermore, there is currently no observational proof that lognormalities on quasi-linear scales do not affect the ISW signal, so there is no motivation to formally exclude these effects from our analysis.

It might occur however that the proposed method projects the statistics of the 2MASS galaxy density field into the phi reconstruction, and, as the authors claim, there may be significant non-Gaussian signal in the 2MASS survey on large angular scales. Since 2MASS is relatively shallow, there may be slight non-Gaussianities (or "log-normalities") projected on relatively large angular scales which could be reflected in the phi prediction performed by the authors' method. But this non-Gaussian component should not give rise to any ISW. One possible way to test if the authors' method indeed project the statistics of the density field into the potential field is by simply applying their method on a pure Gaussian density realization and test the statistics of the inferred potential field.

Our method does indeed use the 2MASS density field to infer the ϕ field. Since there is a one-to-one mapping of the density and ϕ fields in Fourier space, any quasi-linearities in the density field should lead to quasi-linearities in the ϕ field, which in turn will affect the structure of the ISW field (see answer to previous section). In spherical harmonic space, these quasi-linear effects may in practise be negligible.

To test this, we applied our method on both Gaussian and lognormal fields for full sky simulations (i.e. no missing data) to test whether the statistics of the density field will affect the detection level, as suggested by the referee. We found no difference in terms of detection level (see the updated Table 3), apart from a slight difference for the simple correlation method (we suspect this is related to the wild bootstrap method which is more efficient in the Gaussian case).

Furthermore, our detection method does not use any priors on the statistics of the galaxy (or ISW) field and we expect the result to be independent of such information. This discussion has been added to the paper.

I would be happy if that were the case and the non-Gaussianities in the reconstructed phi is due to non-Gaussianities in the 2MASS density field present on the moderate

*to large scales. But this should be then clearly said in the text, and the readers should be made aware that ISW is still associated *only* to the linear (and hence Gaussian for most models) component of the potential.*

Indeed, the non-Gaussianities which may be present in the reconstructed phi field, directly arise from non-Gaussianities which may be found in the density field. As we have discussed above, since the density and phi fields are directly related, it is theoretically possible for the ISW field to present non-Gaussianities. However, the test performed in the previous paragraph and suggested by the referee (see the updated Table 3 for 2MASS and new Table 4 for Euclid) shows that whether the underlying 2MASS or Euclid fields are Gaussian or lognormal has very little effect on detection levels.

In practise, separating the Gaussian and lognormal components of a galaxy or density field is technically very difficult (especially at large scales where few modes are available).

We do however clarify in the text, that we are looking at the ISW signal (i.e. the temperature component which is positively correlated with the matter distribution), and not the non-linear Rees-Sciama effect. We have also added references to Cai et al 2010 and Schaefer et al 2010 which investigate the scale on which the Rees-Sciama effect intervenes.

*Regarding the resulting pdf of the statistic used to measure the cross-correlation, I agree that if both density and potential vary, then the statistic is the product of two normal distributions that is far from a normal one. However, when computing statistical significances, one of them (usually the field inferred from the galaxy survey) is assumed *fixed*. In this case, if one of the two fields (either rho or phi) is fixed, then the pdf will depend on the particular definition of the statistic. I agree with the authors that with their definition of their statistic ($\tau = \rho/\text{Var}(\rho)$) the pdf should be a Student's t-distribution and hence non-Gaussian a priori. But, in my opinion, this would be the case *only* if $\text{Var}(\rho)$ is estimated from the data. If $\text{Var}(\rho)$ can be estimated from theory (i.e., it is a fixed quantity for all MC simulations for a given cosmology) then $\text{Var}(\rho)$ is merely a constant and we should be again in the situation where τ is a Gaussian random variable. For instance, in Eq.(14) the quantity $\sigma_{\bar{\chi}}$ is a quantity that is computed theoretically, and also a theoretical expression can be written for the σ_{ρ} defined in the denominator of Eq.(13). Again, this would only apply if one of the fields (i.e., the one inferred from the galaxy survey) is fixed. I think that the reader would appreciate becoming aware of these issues. In particular, it should be clearly stated which of all cases considered in this discussion was actually used when building the histogram in Fig.1.*

We agree with the reviewer that if the variance is known or estimated independently of the data, then τ follows a Gaussian distribution. We add this remark in the paper.

Regarding Appendix B.2., the bootstrap approach can be more easily understood as it reads now. It is claimed that Davidson & Flachaire (08) proved that this distribution gives exact results in specific cases, but it is not specified which those are. I think that the issue is whether this distribution is able (or not) to provide the correct pdf for the cross-correlation estimates in the particular ISW - LSS scenario (which provides a clear model against which the performance of this approach can be tested). That is, under the usual ideal scenario of Gaussian ISW and LSS fields

it can be computed what the pdf for ρ (in Eq.(B.2.)) should be, and this can be compared to the outcome of this bootstrap approach. I leave this exercise up to the authors, but I would recommend to emphasize in the text that this approach should be taken with caution too, as long as it is not compared to theoretical expectations.

We add the assumptions needed for the claim. We understand the skepticism of the reviewer on the reliability of bootstrap methods, however we think that some points must be highlighted: 1) the bootstrap is only used for the simple correlation test, 2) bootstrap methods allow us to perform tests without knowing the underlying distributions and 3) these methods are known to perform better than asymptotic testing (e.g. most of χ^2 -test) [1]. The bootstrap method has already been successfully used to estimate the confidence intervals in many situations [3, 2, 7] and its reliability depends mainly on the way new samples are generated and on the observation size.

I agree with what the authors say about the difficulties on reconstructing the ISW / potential field out of a survey like 2MASS, and actually I think it should be added in the text, close to Eq.(20). However, the authors have made no comment on my remark on Eq.(20): this Equation is, at best, an approximation for only a part of the $a_{\ell,m}^{ISW}$. One way to see why is that, since it makes no use of the redshift distribution of the sources, but it uses integrated along the line of sight quantities, it can not perform as well as a full 3D inversion of the density field into the potential field. I think the authors should warn the readers about this issue in the text too.

We agree with the referee that the method we use will only lead to an approximation of the ISW field. Of course, the reconstruction of the ISW field using the 3D matter density is the most efficient and precise way to proceed. However, there are currently no full sky spectroscopic galaxy or lensing maps from which we could infer a 3D matter distribution. The ISW reconstruction is still an open problem with some big issues like the missing data and the redshift estimation. This discussion has been added to the paper.

2 Part II, (second level of concern)

I refer again to the argument given above. The ISW (as a linear effect) is due to the linear (and supposedly Gaussian) component of the density field. Whatever 2nd-order (and hence non Gaussian) component present in the density field should not positively correlate with the ISW component of the CMB (but rather anti-correlate to the CMB - i.e., Rees-Sciama effect). Third order effects should be even smaller. Studies of numerical N-body simulations (Cui et al. 08, Smith et al.08, Cui et al. 09, Diego et al. 10) show that the 2nd-order Rees Sciama effect are already negligible. This means that the impact of those log-normalities in the density distribution on the CMB/ISW are hence negligible as well. The ISW should therefore provide no hint whatsoever about non-Gaussianities, provided the linear density field is Gaussian.

If a density peak has non-Gaussian (or log-normal) patterns, these should leave no signature in the ISW component generated by that overdensity, but only on the RS component which is found in simulations to be much smaller, limited to much higher multipoles (and much deeper buried in "cosmological" intrinsic CMB noise).

We have clarified in the text that when we refer to ‘ISW’ signal, we mean the secondary temperature anisotropy which is positively correlated with the density field (but which may have some non-Gaussian signatures, but tends to Gaussianity on the largest scales), and the ‘Rees-Sciama’ effect, as the secondary temperature anisotropy which is negatively correlated with the density field (and is necessarily non-Gaussian).

The detection method which we have developed in this paper is very general and makes only one major assumption: that the primordial CMB behaves as a Gaussian field. The method is also general as it *allows* the galaxy field to behave as a lognormal field, but does not automatically assume that the galaxy field is lognormal (e.g., it can be perfectly Gaussian, and our method allows for this).

I agree with the authors that by using an estimate of the phi/ISW from 2MASS one is focusing the analyses on the angular scales where the S/N is located. However, my request was different: given the angular coverage and the depth of 2MASS, I would like to see the level of S/N that one would expect for a 2MASS-like survey, even in the case where the authors’ method is applied (and not only a simple density x CMB cross correlation approach). Apparently, this is shown in Table 3, where the number of sigmas to be obtained by each method are given after running a number of MC simulations.

The expected ratio S/N can be defined, in our case, as the expected mean detection level we compute through the simulation. For 2MASS we have then a $S/N \approx 1\sigma$ which agrees with previous results [8, 6]. This was clarified inside the section on the expected level of detections.

*I must say this table looks rather odd to me: (1) Numbers quoting # of sigma look a bit too high (for a survey as shallow as 2MASS [$z_{max} \sim 0.15$] a simple comparison with Fig.3 of Frommert et al. [08] show that, under full sky coverage, a perfect density survey up to redshift $z \sim 0.2$ should yield $S/N \approx 1$); (2) the table values referred to the cases with no ISW are of the order of one, and I would have expected them to be close to zero, since if there is no ISW there is no cross correlation between the fields inferred from the galaxy survey and the CMB maps – this is actually a particularly puzzling result: if there is no ISW, we expect *no* cross-correlation whatsoever between the CMB and the galaxy template, I just cannot understand those values quoted in the table–; and (3) similar significance is reached for the inpainted case as compared to the full sky case, for the with-ISW cases (two leftmost columns).*

The confidence of a null hypothesis test can be interpreted as the distance from the data to the null hypothesis (commonly named H_0). For example, let ρ be a variable of interest (e.g. correlation coefficient, amplitude). The confidence score σ for the hypothesis test H_0 (i.e., $\rho = 0$) against H_1 (i.e., $\rho \neq 0$) is directly computed using the formula $\tau = \rho/\sigma(\rho)$, where $\sigma(\rho)$ is the standard deviation. But this is only true when 1) ρ is Gaussian, 2) $\sigma(\rho)$ is computed independently of the observation and 3) considering a symmetric test. Then, this method does not stand for the general case and as the correlation coefficient considered here is positive, an asymmetric test (one-sided test) would also be more appropriate here.

Remember that a confidence score is directly linked to the deviation from the H_0 hypothesis through the p -value which is a probability, so the σ -score is always positive. Then the p -value p of one hypothesis test is computed using the probability density function (PDF) of

the test distribution: $p = 1 - \int_{-\infty}^{\tau} P(x|H_0)dx = \int_{\tau}^{+\infty} P(x|H_0)dx$ (for a classical one-sided test). Remark that in that case, if the H_0 is true then $\rho \approx 0$ (i.e. in the middle of the test distribution) and the p -value p will be around 0.5 which correspond to a confidence of 0.67σ . This paragraph was added to the paper as a subsection.

I would have expected that, since there is no information from the in painted area, the total significance should consequently decrease. It is very hard for me to believe that after inpainting 30% of the sky practically no significance of the cross-correlation is lost. (What would then the achieved significance be if 50% of the sky is first masked and then inpainted?)

It is true that missing data leads to a loss of correlation between the two maps, and that the ratio of missing data will highly impact the success of detecting the correlation. The inpainting method helps us to reduced the variance of the detection, but we can see from the Euclid simulation that we cannot recover the full correlation. This fact was imperceptible with the 2MASS simulation because of the low detection level.

I must confess I am very skeptical about the results of this table, and since they are critical since many of the results of the paper are based upon them, I need to be convinced on their correctness.

We believe that most of this skepticism comes from the way of reading the σ scores. We hope that our explanation above makes the correctness of the method clearer.

I appreciate the effort put in describing the ROC computation. As I understand, simulations are run, and they are sorted according to the p -value provided by each statistic. The Y coordinate is computed then as the ratio of true to false detections. One last thing: how is a "false" or "true" detection defined in this context?

The X-axis corresponds to the false positive rate, i.e. the ratio of CMB maps without ISW where ISW signal is detected at the current threshold. The Y-axis corresponds to the true positive rate, i.e. the ratio of CMB maps with ISW where ISW signal is detected at the current. We recall that a point on the ROC curve corresponds to a threshold. We complete the paper to make it clearer. We have also added the ROC curve for the Euclid survey in the paper.

As they authors say it is hard to decide which of the two maps (ILC and GMCA) are closer to reality. The fact that they yield so different results (2.1 to 2.7 σ) should hence be taken very cautiously. And for this reason I would explicitly write this caveat in the abstract, lowering the tone of the claims (in particular the one pointing to 1.6% probability of false positive).

As we have discussed at the beginning of this answer, we have decided not to include the GMCA data in this paper, as we believe a more thorough analysis of the component separation is necessary, which we plan to do in a further study.

3 Part III, (third level of concern)

I am OK with all points but one: I cannot see why there is no difference when considering $l=2$ (quadrupole). For $l=2$ the ISW has its highest amplitude, and it should also contribute to the ISW total signal.

We agree with the referee that the quadrupole should contribute to the detection of the ISW. We first thought that the quadrupole has no importance in the detection, because with the GMCA map, we hardly see one. However, when we further investigated this using the other maps, we actually find that including or not the quadrupole leads to very different results. The results are actually counter-intuitive, in that including the quadrupole reduces the significance. We suppose that this phenomena is linked with the anomalously low power of the quadrupole.

References

- [1] A. Davidson and J. G. MacKinnon. The power of bootstrap and asymptotic tests. *Journal of Econometrics*, 133:421–441, 2006.
- [2] T. J. DiCiccio and B. Efron. Bootstrap Confidence Intervals. *Statistical Science*, 11(3):189–228, 1996.
- [3] B. Efron. *The Jackknife, the Bootstrap, and Other Resampling Plans*. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial Mathematics, 1987.
- [4] C. L. Francis and J. A. Peacock. An estimate of the local integrated Sachs-Wolfe signal and its impact on cosmic microwave background anomalies. *MNRAS*, 406:14–21, July 2010.
- [5] C. L. Francis and J. A. Peacock. Integrated Sachs-Wolfe measurements with photometric redshift surveys: 2MASS results and future prospects. *MNRAS*, 406:2–13, July 2010.
- [6] T. Giannantonio, R. Scranton, R. G. Crittenden, R. C. Nichol, S. P. Boughn, A. D. Myers, and G. T. Richards. Combined analysis of the integrated Sachs-Wolfe effect and cosmological implications. *Phys. Rev. D*, 77(12):123520–+, June 2008.
- [7] P. Hall. *The Bootstrap and Edgeworth Expansion*. Springer Series in Statistics. Springer, 1995.
- [8] S. Ho, C. Hirata, N. Padmanabhan, U. Seljak, and N. Bahcall. Correlation of CMB with large-scale structure. I. Integrated Sachs-Wolfe tomography and cosmological implications. *Phys. Rev. D*, 78(4):043519–+, Aug. 2008.