Residual foreground contamination in the WMAP data

P Chingangbam$^1$ and C Park$^2$

$^1$ Indian Institute of Astrophysics, Koramangala II Block, Bangalore 560034, India
$^2$ Korea Institute for Advanced Study, Hoegiro 85, Dongdaemun-Gu, Seoul 130-722, Korea
E-mail: prava@iiap.res.in, cbp@kias.re.kr

Abstract. We have studied whether there is any residual foreground contamination in the foreground reduced WMAP-7 data for the differential assemblies (DAs) Q, V and W. We have calculated the correlation between the foreground map, from which long wavelength correlations have been subtracted, and the foreground reduced map for each DA. We have found positive correlations for all the channels. The statistical significance of the resulting values has been tested by comparing with correlations between the cleaned CMB maps and 1000 simulated Gaussian maps to which instrumental effects have been added. We have found high statistical significance of the observed correlations, implying the presence of residual contamination in the cleaned data, and found that, for Q and V channels, a large fraction of the contamination comes from pixels, where the foreground maps have positive values larger than three times its rms value, which shows the presence of unresolved point sources that contribute significantly to the contamination.

1. Introduction
The experimentally measured CMB temperature fluctuations is composed of the true CMB signal and foreground contamination, coming from astrophysical sources that emit photons in the frequency ranges spanned by the experiments. The major part of the contamination comes from emissions in our galaxy, and a small fraction comes from point sources such as radio galaxies and dusty star-forming galaxies. It is necessary to know the contaminating component in order to get the true signal, which can be used to extract cosmological information [1]. The galactic foreground is usually estimated from a combination of theoretical modelling and observations, and then subtracted from the measured data [2]. Further, sky regions where there are possibility of residual contamination, such as the directions along the plane of our galaxy, and the locations of localized sources that have been identified are masked. The correct removal of the foreground contamination is extremely crucial, because any residual contamination will bias the extraction of cosmological information from CMB data.

In this paper, we have investigated whether there is small residual foreground contamination in the cleaned WMAP data by using a method that is simple, but has proved to be quite effective. Our method is based on calculating correlations between the foreground field, which has been processed so as to remove long wavelength correlation of the galaxy emissions, and the cleaned CMB data. Our basic premise is that if there is no residual contamination in the cleaned data, we should obtain no correlation. However, we have found statistically significant positive
correlation for WMAP-7 data for the Q, V and W differential assemblies (DAs)

Further, we have found that a big fraction (as big as 30% for Q channel) comes from regions where the foreground map has large positive values, which indicates unresolved point sources. This is not surprising, in light of the fact that the PLANCK early data release has identified point sources of the order of 15000 [3]. These results imply the presence of residual contamination in the cleaned data. This work has been done with the motivation to quantify the contaminant fraction, and understand the nature of non-Gaussian deviations they introduce into the CMB. The earlier studies on point sources contamination on the CMB have been included [4, 5, 6, 7].

This paper is organized as follows: In Section 2, we have outlined our procedure for calculating the correlations between the foreground and cleaned CMB maps, and then presented our correlation results and their statistical significance. We have given the concluding remarks in Section 3.

2. Quantifying residual foreground contamination

Let $f_1$ and $f_2$ be any two random fluctuation fields having zero mean values, defined on the surface of a two dimensional sphere. Let their rms values be given by $\sigma_1$ and $\sigma_2$ respectively. Re-scaling them as $\nu_1(i) \equiv f_1(i)/\sigma_1$ and $\nu_2(i) \equiv f_2(i)/\sigma_2$, where $i$ denotes the pixel number, we can define a correlation parameter $r$ as:

$$r \equiv \langle \nu_1(i) \nu_2(i) \rangle,$$

where the bracket denotes average over all pixels. We expect $r$ to be zero if the two fields are uncorrelated, and non-zero otherwise. (In practice, we will always get a small, but non-zero value of $r$ even though the two fields are known to be totally uncorrelated, due to the finite number of pixels.)

The observed WMAP data $f^{\text{obs}}$ is a sum of the true CMB signal and foreground contamination. The foreground component is estimated using a combination of galaxy observation and theoretical modelling [2]. We have called this field the ‘apparent’ foreground field, denoted by $f^{\text{appfg}}$, keeping in mind that there may be small error in its estimation. This field is then subtracted pixel by pixel from $f^{\text{obs}}$ to leave behind the ‘cleaned’ CMB signal, which we have denoted by $f^{\text{cleaned}}$. By definition, $f^{\text{cleaned}}$ has zero mean. If $f^{\text{appfg}}$ has been correctly estimated, then we expect it to have negligibly small correlation with $f^{\text{cleaned}}$, since they come from totally different physical processes. However, if the estimation is on the right track but not fully correct, then we should expect some residual contamination in the signal field. This should show up as non-zero correlation between $f^{\text{cleaned}}$ and $f^{\text{appfg}}$. In what follows, we have described our results for the correlations.

2.1. Peak field

Our analysis is done using the 7 years data from the eight DAs of WMAP [8], namely, Q1, Q2, V1, V2, W1, W2, W3, and W4. For each DA, we have obtained $f^{\text{appfg}}$ as:

$$f^{\text{appfg}} = f^{\text{obs}} - f^{\text{cleaned}},$$

Since, we want to remove the long distance correlations of the emissions from our galaxy, we have first defined what we call the peak field as:

$$f^{\text{peak}} \equiv \left(f^{\text{appfg}, \theta_s} - f^{\text{appfg}, 3\theta_s}\right) - \left(f^{\text{appfg}, \theta_s} - f^{\text{appfg}, 3\theta_s}\right),$$

where $\theta_s$ and $3\theta_s$ are FWHM values, at which we have smoothened the field. By definition, $f^{\text{peak}}$ has zero mean. Figure 1 shows the peak field for Q1 channel.

1 We have done similar analysis using WMAP-5 data and obtained similar results.
Let us denote:

\[ r_c = \frac{\nu_{\text{peak}}}{\sigma_{\text{peak}}} \]

and define:

\[ r_c \equiv (\nu_{\text{cleaned}}, \nu_{\text{peak}})_{\theta_s} \]

where the suffix \( \theta_s \) is to remind us that we do the calculation for a choice of FWHM at which \( f_{\text{cleaned}} \) has also been smoothened. \( \sigma_{\text{cleaned}} \) and \( \sigma_{\text{peak}} \) are of the orders of \( 10^{-5} \) and \( 10^{-7} \) respectively.

Table 1 summarizes the main results for \( r_c \). We have used \( \theta_s = 35' \). Two values of \( r_c \) are shown for each DA and \( s_{\text{mask}} \). The upper value is the case where \( r_c \) is calculated using all unmasked pixels, while the lower value is the case where pixels with \( \nu_{\text{peak}} > 3 \) have also been excluded. The first observation we have made is that all \( r_c \) values are positive. For Q channels,
we get considerably larger correlation when we keep all unmasked pixels, larger by about 30%. This indicates that there is non-trivial correlation arising from the pixels with $\nu_{\text{peak}} > 3$. For V channels, the difference is about 20%, while W channels do not seem to be affected. We have repeated these calculations for the case where $L_{\text{cleaned}}$ are also excluded, but not found any significant effect. The sky fractions for the three $s_{\text{mask}}$ values are roughly, 62%, 60% and 58% respectively. For Q and V channels, as we stay further away from the mask boundaries, there is small but systematic decrease of $r_c$.

2.3. Statistical significance of $r_c$ values

We have investigated how likely it is to get the observed $r_c$ values given above by comparing with correlations between the peak field and Gaussian CMB simulations. For this purpose, we have simulated 1000 Gaussian CMB maps with WMAP-7 parameter values, and added pixel window effect, beam smearing and WMAP-7 noise characteristics. Next, we have smoothened by FWHM 35′ and mask in exactly the same way as we did when calculating $r_c$, and calculated the correlation with the peak field. We have denoted the correlation value by $r_g$. The Gaussian fields are uncorrelated with the signal field, and we should get small value of $r_g$. This exercise has told us what is the typical value of ‘small’ $r_c$ that we can approximate it to be zero for the number of pixels under consideration, and how likely are our observed $r_c$ values to occur by random fluctuation and not due to a true correlation.

We have counted, out of the thousand $r_g$ values, how many are greater than $r_c$. The results are shown in Table 2. The left table shows the number $N$, of Gaussian maps having $r_g > r_c$ for each individual DA, for the three $s_{\text{mask}}$ values used earlier for calculating $r_c$, and including/excluding the pixels having $\nu_{\text{peak}} > 3$. When all unmasked pixels are included, we get $N = 0$ for all $s_{\text{mask}}$ values for Q channel, for V channels, $N$ lies between 5 and 15, while for W channels, $N$ lies between 121 and 291. These numbers imply that the $r_c$ values for Q and V are statistically significant, whereas, the values for W channels have much lower significance. When pixels with $\nu_{\text{peak}} > 3$ are also excluded, we have found a reduction of $N$ for all the channels. The table on the right side of Table 2 shows the number $N_0$, of Gaussian maps having $r_g > r_c$ simultaneously for all DAs. These values are again significant. Therefore, we have concluded that the cleaned WMAP data, particularly Q and V channels, contain small but statistically significant amount of residual foreground contamination.

Table 2. Left: Number of maps $N$, having $r_g > r_c$ for individual DAs, out of 1000 Gaussian maps. As in Table 1, upper values are for all unmasked pixels included, while lower values are for the case when pixels with $\nu_{\text{peak}} > 3$ have also been excluded. Right: Number of Gaussian maps $N_0$, having $r_g > r_c$ simultaneously for all DAs.
3. Conclusion
We have shown that the cleaned WMAP data contains small but statistically significant amount of residual foreground contamination. The level of contamination is quantified by the level of correlation obtained between the cleaned map and the peak field. For Q and V DAs, we have found that a big fraction of the contamination could be coming from unresolved point sources. Our result has important implications for the extraction of cosmological data using the cleaned WMAP data. For example, it is known that constraints on the local primordial non-Gaussianity parameter $f_{NL}$ using Fourier space method such as the bispectrum, and those use pixel space methods, like the Minkowski Functionals (MFs) do not agree with each other. The bispectrum gives positive best fit value of $f_{NL}$ [9], while the MFs give negative values [10]. Our analysis in this work was motivated by an attempt to understand whether this conflict is due to residual foreground contamination. Our next goal is to understand the effect of the residual contamination on the measurement of the MFs, and we are currently pursuing this line of investigation.

Acknowledgement
We thank Korea Institute for Advanced Study for providing computing resources (KIAS Center for Advanced Computation Linux Cluster System QUEST), where a part of the computation was carried out. We acknowledge the use of the HEALPIX package [11, 12], and the Legacy Archive for Microwave Background Data Analysis (LAMBDA). Support for LAMBDA is provided by the NASA Office of Space Science.

References
[1] Komatsu E et al 2011 Ap. J. Supp. 192 18
[2] Gold B et al 2011 Ap. J. Supp. 192 15
[3] Planck early results VII. The Early Release Compact Source Catalogue 2001 A. & A. 536 A7
[4] Argueso F, Gonzalez-Nuevo J and Toffolatti L 2003 Ap. J. 598 86
[5] Boughn S P and Partridge R B 2008 PASP 120 No. 865
[6] Babich D and Pierpaoli E 2008 Phys. Rev. D 77 123011
[7] Lacasa F, Aghanim N, Kunz M and Frommert M [arXiv:1107.2251[astro-ph.CO]]
[8] http://lambda.gsfc.nasa.gov/
[9] Komatsu E et al (WMAP Collaboration) 2009 Ap. J. Supp. 180 330
[10] Hikage C, Matsubara T, Coles P, Liguori M, Hansen F K and Matarrese S 2008 MNRAS 389 1439
[11] Gorski K M, Hivon E, Banday A J, Wandelt B B, Hansen F K, Reinecke M and Bartelmann M 2005 Ap. J. 622 759 [arXiv:astro-ph/0409513]
[12] http://healpix.jpl.nasa.gov