Dear referee,
Thank you for your comments and corrections, which will improve the quality of this work.
Please find below the answer to your report. The changes in the text have been made in bold face.
Best regards.
Rémi Adam, Iacopo Bartalucci, Gabriel Pratt et al.

The present paper attempts to extract the first resolved map of the kinetic Sunyaev-Zel'dovich (kSZ) effect toward a galaxy cluster with the NIKA camera. The authors present the kSZ signal map of MACS J0717.5+3745 at z=0.55 and discuss that the system is undergoing a subcluster merger. I however have several concerns about their results as described below. The following points should be taken into account properly before the paper is considered for publication further.

1) A high significance detection of the kSZ signal is reported only toward region B. If it is real, the strong negative signal at 260 GHz in Figure 1 would be surprising given the high electron density expected from the X-ray data toward this region, even if one allows for uncertainties in the gas distribution along the line-of-sight. The authors should provide further justification for the detection by
a) Presenting the noise maps from the difference between two equivalent subsamples (without normalizing each pixel) described in Section 2.2 and testing if the noise is Gaussian
b) Giving details on how the zero brightness level of the NIKA maps was determined and taking account of its systematic errors
c) Quantifying the degree of primary CMB contamination over the field-of-view (not only at the 22" resolution)

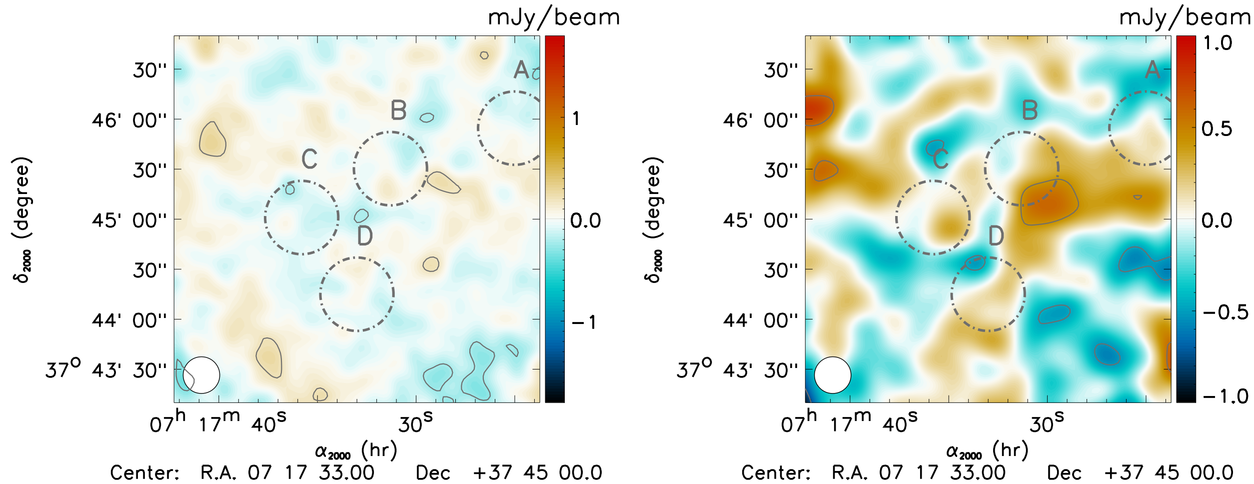
It is true that a high density in region B would induce positive tSZ signal at 260 GHz. However, the NIKA 260 GHz band is relatively close to the zero of the tSZ spectrum and close to the maximum of the kSZ spectrum (the sensitivity of the NIKA band to kSZ and tSZ are given in Table 1). A high electron density in region B would therefore also result in stronger kSZ signal, and thus more negative brightness at 260 GHz, in the case of positive line-of-sight velocity. As the tSZ signal is always positive at 260 GHz, and seems to be canceled by kSZ signal, this means either that the moving gas is very dense, but very cold, or that it is moving very fast (or both). In case of high electron density, the fact that the 150 GHz signal is relatively low suggests that the gas is relatively cold.

In addition, we stress that the filtering of the NIKA processing removes the diffuse signal on large scales while keeping compact signal. The tSZ signal being a priori less clumpy than the kSZ one, we expect to pick-up more easily the sub-structures in the signal it induces.

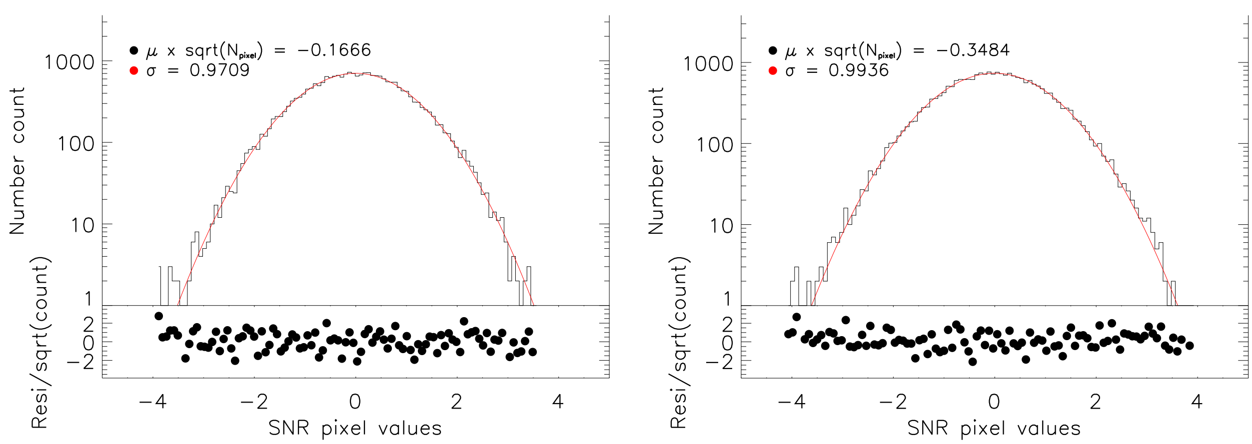
For these reasons, we are not surprised to observe negative signal at 260 GHz (with relatively low significance though) even if it is indeed slightly larger than what we could expect from X-ray data and simple cluster geometry. Moreover, we are confident about the reliability of our data, within the quoted uncertainties, as we show in more details below.

<u>a) Jackknife maps and gaussianity</u>

The maps presented below are the same as figure 1, in the case of half difference maps of equivalent subsamples (thus at 22 arcsec FWHM resolution). Contours are multiples of 1 sigma in both cases, starting at +/-2.
This figure has been added to the paper.



The noise gaussianity is indeed an assumption of our procedure. It was checked in previous works, but was not published. The figure below provides the signal-to-noise ratio histogram (without smoothing in this case) and its Gaussian best fit. We do not observe significant deviation from gaussianity and the mean and variance are in agreement with expectations.
We now mention that the gaussianity of the noise was checked, but we do not think it is necessary to add this figures in the paper.

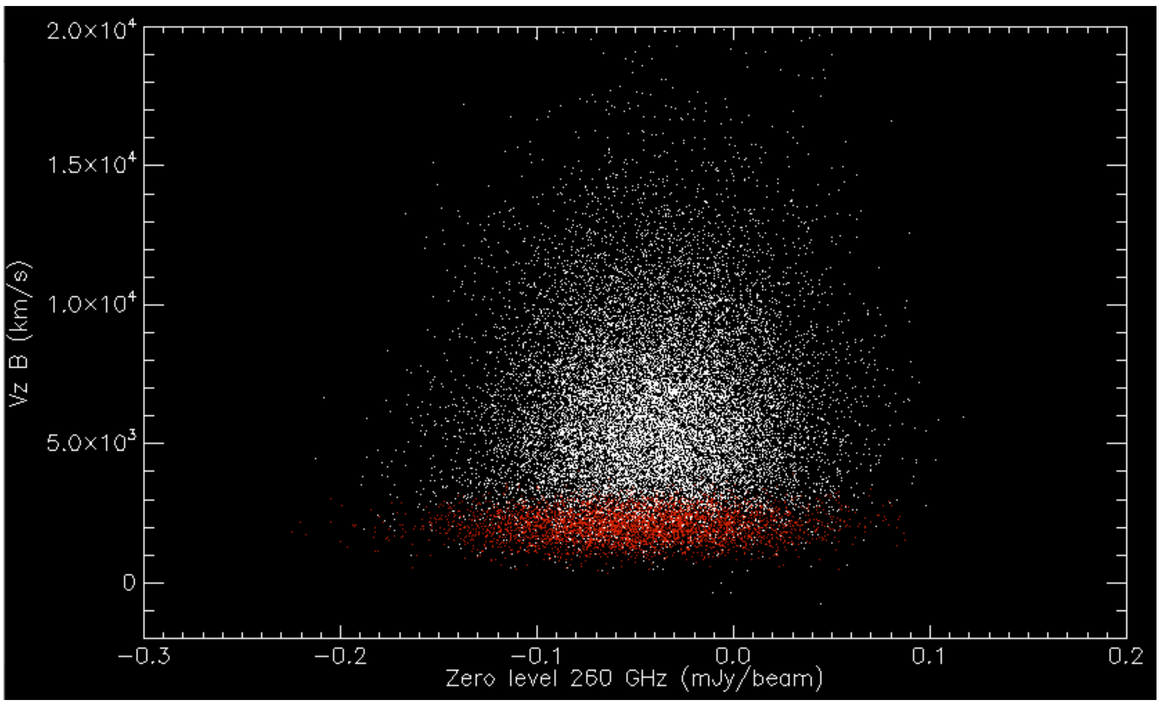


b) Zero level effects

When reducing the data, in the mapmaking procedure, the zero level is computed by setting the mean of the time ordered data to zero, taken in regions that are not flagged as on-cluster (in an iterative manner). The flagged region is based on the 150 GHz map because of the higher signal to noise.

1 - kSZ map (section 5): In the case of the kSZ map, this means that the average signal is set to zero in the external parts of the cluster (based on the 150 GHz image). In fact, the transfer function of the NIKA reduction is affecting the kSZ map the same way it affects the 150 and 260 GHz maps.

2 - Velocity fit (section 6): When we model the cluster and fit for the gas velocity, we marginalize over the zero level of the maps. This means that we also include the zero levels in the model as nuisance parameters, which are fitted independently for each map, without any prior. In practice, the zero levels are not correlated with the velocity. Indeed, the kSZ map is mainly bimodal, with no structures on large scales. Therefore, changing the zero level does not improve the chi2, i.e., it is not possible to mimic the kSZ signal that is observed by changing the zero level. In the case of an isolated cluster moving along the line of sight, the zero level would probably be correlated with the kSZ signal fitting parameters. For illustration, the figure below shows the MCMC chain samples for model F1 (white) and F2 (red) in the plane zero level - velocity.
More details have been provided in the text for both cases, and in section 2.2.



c) CMB contamination

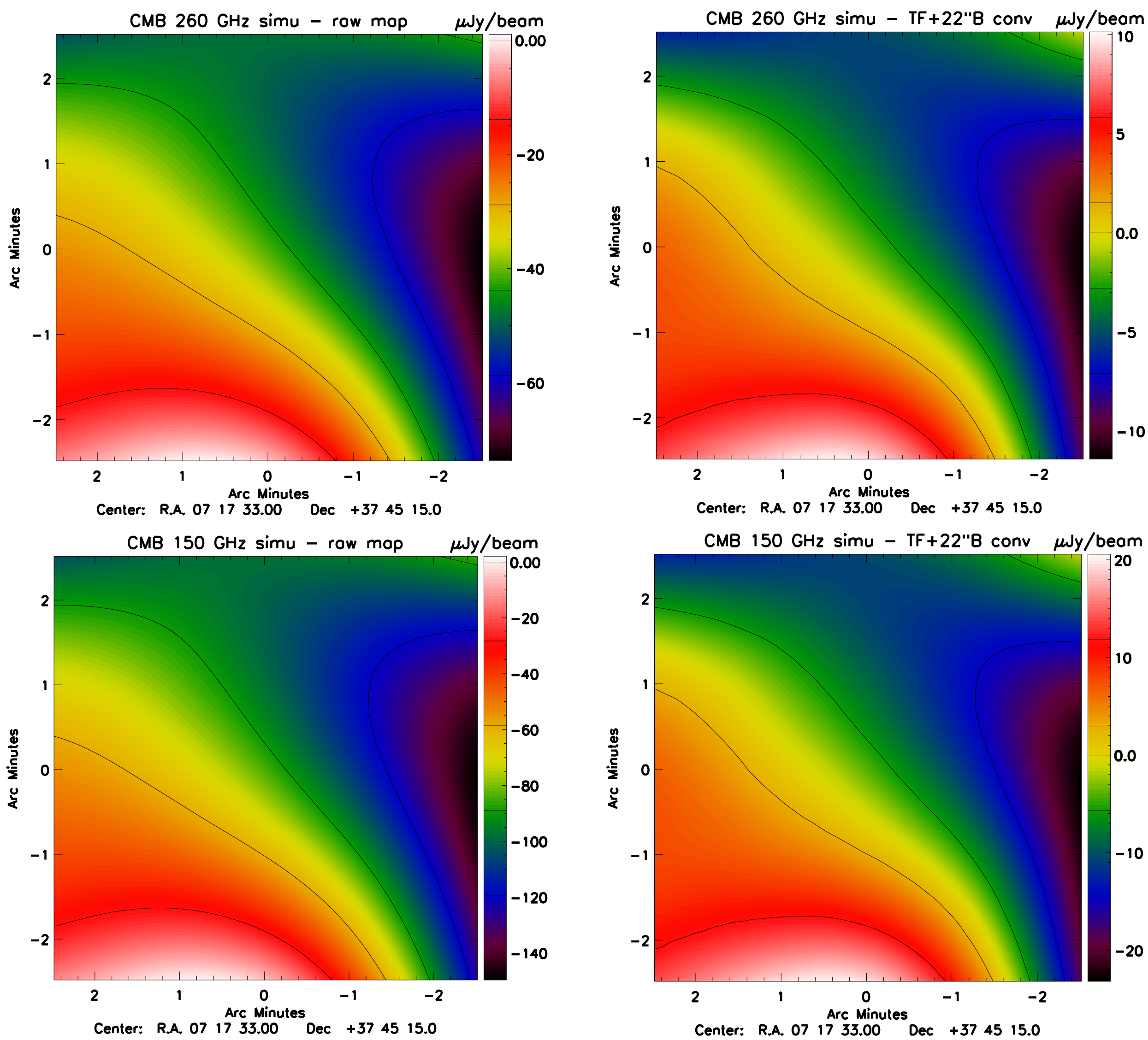
The CMB angular power spectrum being very steep at the considered angular scales, the 22" beam smoothing effects are negligible. In fact, the large scale filtering of the transfer function has much more impact on its expected signal. However, even ignoring the transfer function, the CMB would be negligible at the scale of our field-of-view.

The figures below are examples of a CAMB CMB realization at both 150 and 260 GHz in the field of view of the NIKA MACS J0717.5+3745 data. It is shown in the case of raw CMB maps and in the case where the transfer function (including the 22" beam smoothing) has been applied to the maps. The CMB induced noise rms taken in this 5x5 arcmin field is:

31 uJy/beam for the raw 260 GHz map
8 uJy/beam for the transfer function filtered 260 GHz map
64 uJy/beam for the raw 150 GHz map
16 uJy/beam for the transfer function filtered 150 GHz map

The rms taken at the scales of the observed cluster sub-structures is even much smaller.

The CMB contamination was already discussed and quantified in Adam et. al 2016 (arxiv:1510.06674), as cited in the present paper. Therefore, we do not think it is necessary to include such figures or more discussion in the paper.



2) Figure 7 indicates that the 260 GHz data toward the kSZ peak in region B disagree at more than 2-sigma with any models of the line-of-sight velocity and the electron density considered in the paper. I suppose that the disagreement becomes even larger once the X-ray imaging data are included in the analysis. The authors should present a counterpart of Figure 7 for F2 and discuss implications of the disagreement in detail. Note also that the quoted significance of the fitted velocity (e.g., 5.1 sigma for region B) is not meaningful if the fit itself is statistically unacceptable.

Indeed, we observe a -2.3 sigma peak on the residual toward region B at 260 GHz in case of model F1, which becomes -3.1 sigma in case of F2. This means that the overall best fit, including the two NIKA maps, fails to completely describe the negative peak at 260 GHz in region B. Some models provide a better fit at 260 GHz in region B, but the residual increases at 150 GHz for them. This leads to an overall poorer chi2 since the SNR is much higher at 150 GHz, even though we also observe other residual features at 150 GHz (more than statistically allowed). Considering the complexity of MACS J0717.5+3745, it is not surprising that the model is not perfect. In the case of region B, we think in particular that this may be due to two effects: 1) the isothermal assumption of this model (i.e. in practice the tSZ template should be less peaked than the kSZ one, but it is not the case, and the profile shape is in large part driven by tSZ at 150 GHz); 2) the fact that sub-cluster B is likely made of more than one sub-components itself. These effects could probably be taken into account using a more complex model, but it would be somehow arbitrary (e.g. which shape for the temperature profile?) and it is not strongly justified by our data.

At 260 GHz, because of the lower SNR, the residual is on overall much more consistent with noise than it is at 150 GHz, and the residual is not significantly degraded for F2 with respect to F1. The root mean square of the residual signal to noise is:

0.856 at 260 GHz for F1 and 0.861 for F2
1.181 at 150 GHz for F1 and 1.191 for F2

Concerning the quoted significance values (e.g. 5,1 sigma), we stress that this is not related to the fit or any modeling of the cluster. It corresponds to the peak SNR of figure 5 right as discussed in section 5.

The figure has been added for model F2, and more details have been included concerning the limitation of the model.

Minor comments:

3) It is not clear to what extent the errors associated with the point source subtraction are taken into account in the analysis. The confidence levels presented in the paper, e.g., those in Figure 5, should include such errors.

Fortunately, the possible contamination of the signal from point sources mis-subtraction residuals is local and resolved with NIKA. It is not trivial to propagate the errors from point source residuals for two reasons: 1) the likelihood of the flux of the sources we obtain in the multi-wavelength SED fit are not necessarily Gaussian and we do not know the exact function that describe it. Thus, propagating the full likelihood would require very good likelihood sampling, and therefore very heavy Monte Carlo simulations, 2) it is not guaranteed that the errors on the fluxes estimate are dominated by statistical errors in our fit. In fact the contribution arising from mis-modeling is likely to be important for some sources (e.g. radio source F, as discussed in section 3.2). In practice, this is why we prefer to mask the possible contaminated regions.

In the kSZ mapping analysis (section 5), the point sources are masked (mask of figure 2) when computing the spectra (figure 4). The figure 5 is now updated to show the region of possible point sources mis-subtraction, (i.e. lower confidence) but we do not directly include them in terms of SNR contours for the reason detailed above.

In the velocity fit analysis (section 6), the mask of figure 2 is also used to reject possibly contaminated pixels, so the point sources do not impact the results.

More details have been provided in the paper. Figure 5 now includes the positions of the point sources.

4) I wonder why the confidence contours appear to be identical between the bottom two panels in Figure 9.
The contours are indeed identical for F1 and F2, and they are also identical to the kSZ map SNR contours. While the absolute value of the velocity depends on the optical depth (from which the associated systematic errors are not propagated), the signal-to-noise of the velocity is the same as the one directly measure from the kSZ map (v_z = y_kSZ/tau). The only difference between the two maps of figure 9 is tau.

This is now written more explicitly in the paper.

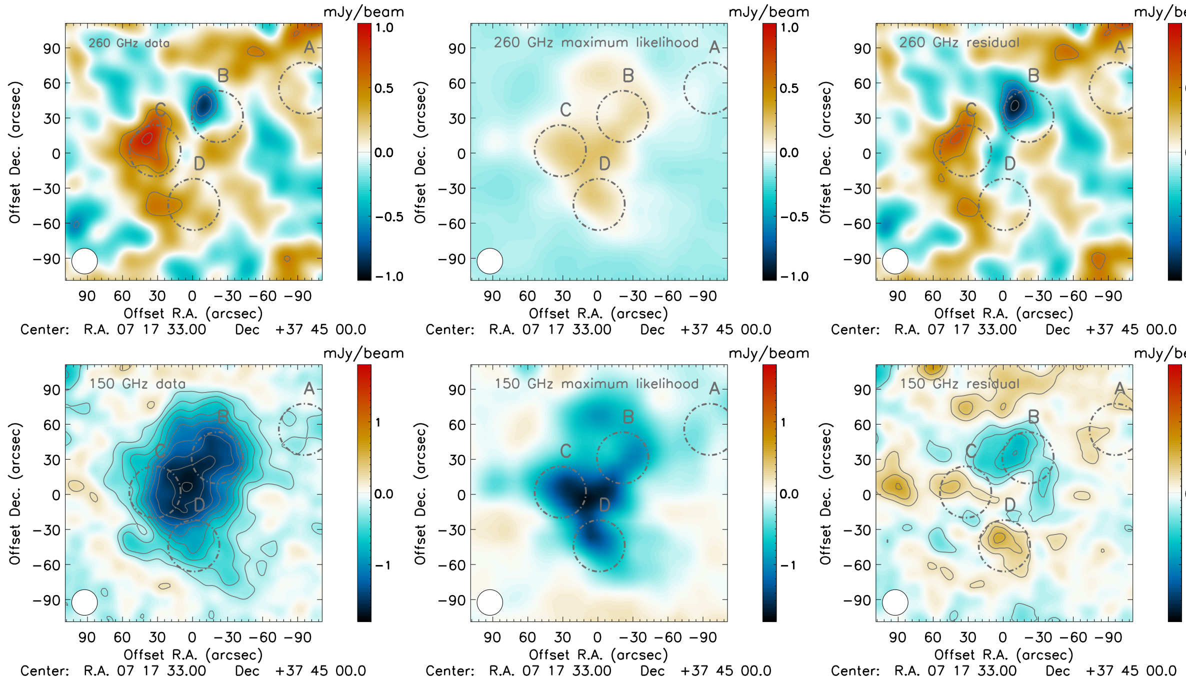
5) Given that a detection of the kSZ signal was reported previously from the sky same region, the authors should perform more direct comparison between their results and those of Sayers et al. (2013) adopting the same effective beam sizes. If the difference of the transfer functions matters, the same gas model can be used for comparison.
Direct comparison between Bolocam and NIKA:

It would be very interesting to compare quantitatively the maps of NIKA and Bolocam, and even combine them for a deeper/higher SNR analysis. NIKA is sensitive to the range [~15" - ~2'] while Bolocam is sensitive to [~45" - ~8'], so the overlap is relatively small, but the complementarity is very good. The bandpasses are also slightly different (the NIKA ones are ~10 GHz higher at 150 GHz and ~10 GHz lower at 260 GHz). However, such combined analysis or detailed comparison is really beyond the scope of the present paper. Indeed, the Bolocam maps used in Sayers et al. (2013) are not publically available to our knowledge, so that such work should be made together with the Bolocam group. We would also need in particular the transfer function, beams and, bandpasses corresponding to the Bolocam maps.
Pseudo Compton parameter map based gas model:

Using the same gas model as in Sayers et al. (2013) is feasible. It was not done at the time of submission because we think that clumping and X-ray temperature uncertainties (especially at this high temperature) are a major limitation of this model at the scales probed by NIKA. They could induce fake signal wrongly interpreted as kSZ, as we discuss below. Nevertheless, we have now performed a similar gas modeling as in Sayers et al. (2013):

1) Construction of a pseudo Compton parameter map using XMM surface brightness and XMM spectroscopic temperature map (see eq 5 of Sayers et al. 2013). We also used the XMM reconstructed density profile to compute an effective electron depth map, ℓ = (integral n_e dl)^2 / integral n_e^2 dl, which is needed in this procedure.
2) We reproduce the fit described in the present paper, but instead of using a tSZ model based on the isothermal beta-models, we fit the tSZ template computed from the pseudo Compton parameter map processed in our transfer function. The kSZ signal remains described by the beta-model.

The figure below illustrates the fit residuals when including only the tSZ templates. We can see several significant features a priori unrelated to kSZ, which might be induced by artifacts in the pseudo Compton parameter template (clumping via the limitation of the electron depth map ℓ, the X-ray temperature uncertainties). Some other features are due to X-ray point sources but they are masked in the fit (e.g. the positive peak on the left/middle).



Concerning the clumping, we stress that the clumping factor, C = <n_e^2>/<n_e>^2, is directly related to the electron depth map, and the pseudo Compton parameter goes as sqrt(ℓ). For mergers, we expect C~1.4-1.6 (Zhuravleva et al. 2013). While a single number is probably sufficient to provide an electron depth at the scales probed by Bolocam, this map cannot account for small scales inhomogeneity to which we are sensitive with NIKA. We thus expect the pseudo Compton parameter map to show local inaccuracies of the order of 20%, which could be responsible for a significant part of the residuals of the figure below.

Comparison to the kSZ map:

Despite the limitations of this model for NIKA, we note that the residuals show that the negative kSZ signal in region B is also very significant from the 150 GHz map only. The position of the peak is the same as the one of the negative peak at 260 GHz (or on the kSZ map) and coincides with sub-cluster B1 (but not B2, see figure 6) reinforcing our multi-components hypothesis for region B. We also observe a positive residual in region C, coincident with the kSZ map, but it is less significant.

Pseudo Compton parameter map based model versus multi-component beta model:

In the pseudo Compton parameter map based gas model, the fit of the kSZ component is driven essentially by differences between the X-ray based tSZ model and the 150 GHz data because the 150 GHz map SNR is much higher than the 260 GHz one. The fit is thus very likely to be affected by the contamination described above. In the case of our multi-component isothermal beta model, the constraint is given by the difference between the two NIKA maps themselves, such that mis-modeling effects can show up on the residuals, but their impacts on the velocity constraint are less important.

As the gas model fit based on the pseudo Compton parameter map appears to be very risky in our case, and since it does not improve the overall gas modeling with respect to our data, we prefer not to include it to avoid adding confusion in the modeling section of the paper. Nonetheless, we provide detailed discussion of our choice and comparison between the two approaches.