nature portfolio

https://doi.org/10.1038/s41586-021-03385-7

Supplementary information

All-sky dynamical response of the Galactic halo to the Large Magellanic Cloud

In the format provided by the
authors and unedited

Peer Review File

Manuscript Title: All-Sky Dynamical Response of the Galactic Halo to the Magellanic Clou

Redactions – Third Party Material

Parts of this Peer Review File have been redacted as indicated to remove third-party material.

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referees' comments:

Referee #1 (Remarks to the Author):

As the title of the paper indicates, Conroy and collaborators claim to have detected "the all-sky response of the Galactic halo to the Magallanic Clouds". The authors compute the spatial distribution of 720 M-giants within 60<r/kpc<80 from the Galactic center and find that it is highly anisotropic. Two over-densities stand out, one in the Southern hemisphere at l>0 (density wake), and another in the Northern hemisphere at l<0 (global effect). Comparison with a numerical simulation that describes the response of the Galactic halo to the accretion of the LMC shows a relatively good agreement with the location of the two over-densities, which the authors use as evidence that these features are caused by a massive LMC falling into the Milky Way for the first time.

I find the topic and the results extremely timely. The results have very strong implications for future modeling of the Milky Way potential and will be debated for many years to come. However, I would like the authors to address a few unclear aspects of the analysis before I can recommend this paper for publication.

- The authors will agree with me that the all-sky response should also be expected in other halo tracers beyond M-giants. Catalogs of K-giants, BHBs and RR Lyrae are publicly available, and I would strongly recommend the authors to analyze them in a similar fashion. This paper would be come _much_ stronger if the authors find a similar features in the distribution of K-giants, BHBs and RR Lyrae. Conversely, the authors may find that the signal only arises in the M-giant population, which would indicate that the anisotropic distribution of stars is not connected to the LMC, or that some flaw exists in the analysis. In my opinion, such a comparison would provide a crucial test of their results.

- Fig. 1 the authors claim good agreement between data and simulation. But I have found quite strong discrepancies
* the model predicts an over-density at l>0, b>0, but the data shows an under-density.
* the amplitude of the global response is Delta rho/<rho>~0.8, but the model predicts Delta rho/<rho>~0.1.
* the amplitude and location of of the wake is Delta rho/<rho>~0.8, but the model predicts Delta rho/<rho>~0.2.

To my surprise, the text literally states that "the simulation reproduces the location and general behavior of the wake to a remarkable degree", which in my opinion is not supported by the results.
I am not sure what to take from this. Either the models are not realistic, or the over-densities are not caused (either entirely or partially) by the LMC. Both alternatives weaken the main conclusion

of the paper.

- How do the results depend on the choice of initial conditions for the stellar halo? is it possible to mimic the anisotropy shown in Fig 1 by choosing a triaxial halo with the right orientation ?

- The analysis relies on a severe distance cut 60<r/kpc<80, but I cannot find any clear justification for this particular choice. Are they using the N-body model to inform them on the distance range to inspect? Are the two over-densities still visible without this interval?

- The coordinates need some clarification. The text claims that the coordinates are Galactocentric but not centered at the sun? Is longitude=0 oriented with the Sun? Or is longitude 180? Would it not be more intuitive to use spherical coordinates phi and pi/2-theta instead of re-definining l and b?

- It would be _extremely_ helpful to add longitude and latitude lines to Fig 1 and Ext Material Figs 3, 4 and 5.

- if the coordinates are centered at the Galaxy, why there is an over-density in Extended Data Fig 4 at l=+- 180 deg.? if the halo is axisymmetric, I would expect no dependence on longitude.

- The analysis assumes that the stellar halo is made of stars with a metalicity Fe/H=-1.5 and an age of 10 Gyr. However, the stellar halo shows a non-negligible spread in metalicities and ages, so adopting delta-functions may bias the results in a non-trivial way. How sensitive is the location/amplitude of the over-densities to the choice of age/metallicity? how would it change if full distributions (e.g. Gaussians) are used instead of delta functions?

- The authors use stars at latitudes |b|>10 degress, but only show photometry of stars at b>45 deg (Ext.data Fig 1). How does the photometry look for stars with 10<|b|<45 ? does the purity of the sample increase or decrease at low latitudes?

- The authors claim that the sample has high purity, but one can see by eye that there is significant contamination at RP-W1<1.8. How much do the result depend on this cut? could the authors show how Fig 1 changes if the cut is placed at RP-W1> 1.9?

- what is the completeness of the sample? can the author guarantee that fraction of M-giants detected is homogeneous across the sky? This can hardly be the case if we consider, for example, extinction. If completeness is not spatially homogeneous, how is this taken into account in the maps constructed in Extended data fig 3?

- When the authors compute the average counts per pixel, do they take into account that a significant fraction of the sky is masked? For example, can the authors exclude that the apparent overdensity at l>0, b>0 simply arises because most of the quadrant containing the MCs has been masked? How is the average density _exactly_ calculated?

- The number count in Extended data fig 3 ranges from 0 to 3. This means that the statistics governing the density maps (prior to smoothing) is deeply Poissonian. How is this taken into account when constructing maps? Also, is the plot in Extended data fig 3 sampling the halo component with Poissonian (low-number) statistics, or are the authors simply using all the particles in the model? To compare apples with apples, my suggestion would be to use the same number of particles as stars in the catalogue (~720) and smooth data and simulations in the same way (including masks) in order to allow a meaningful comparison of Delta rho between models and data.

-The colour bars in Extended Data Figure 3 must be wrong. There are pixels with 3 stars in the right panel that don't have 3 stars in the left panel. Where did these stars come from?

-In general, the description of what they did with the N-body model is not anywhere near robust enough. What assumptions go into building the LMC-MW models? Are they treating stellar halo as a compilation of individual particles or re-weighting DM particles?

- what are the Initial Conditions for the MW stellar halo? could the authors show the radial profile of M-giants and the stellar halo model? The authors claim that they sample the model to have the same radial distribution as the data, but I am not convinced this will give a fair (or unique, at 720 stars) answer. For example, how could the contrast change by sampling from stellar halo models with different (initial) slopes?

- why the distance cut at r>60kpc does not remove all disc stars? does it mean that the sample contains missclassified disk stars as M-giants?

- I am not sure how the proper motion cuts are chosen to remove disk stars. Could the authors show how it works?

Referee #2 (Remarks to the Author):

Please find below my report on "Detection of the All-Sky Response of the Galactic Halo to the Magellanic Clouds", by C. Conroy and collaborators.

This is a very interesting and well-written study, reporting the discovery of the long sought-for response of the Galactic halo due to the arrival of the Magellanic system. Showing the reality of gravitational wakes is of extreme importance for our understanding of the dark matter - indeed it would provide smoking-gun evidence for the particulate nature of this unseen component of our Universe. I am aware of many attempts to try to find the LMC wake (including by my own team!) that did not result in a positive detection. The key insight that is brought to bear on this problem in the present manuscript is shown on Figure 1 in the Extended Data, which illustrates the new photometric parallax technique they have found (and which I have verified). In my opinion, this is an important result in itself.

On the other hand, I am concerned about the reality and statistical significance of the claimed detection of the LMC/SMC wake. I would like to see the following additional analysis and clarifications before recommending publication:

1) While I agree that selecting in the distance interval of 60 to 80 kpc seems like a reasonable choice, what do adjacent distance cuts reveal? From my own selections trying to reproduce the present result, I see that there are many more stars in the 50-60kpc range.

2) I applied the author's selection (but without the complex spatial window) to the public LAMOST DR5 sample. Interestingly, there is a substantial metallcity range of these potential "halo" stars: -1.7 < [Fe/H] < -0.4, I presume due to the presence of sub-structure or contamination. This suggests to me that contamination may be a concern.

3) The smooth halo model that is used as a comparison presumably assumes that the stellar populations in the halo are also perfectly mixed. However, the large number of known halo substructures demonstrates that this assumption is not perfectly valid, and many chemically-different halo features may yet lie just below present detection limits. Given the halo's steep density profile, one may then worry that small deviations from this assumption could lead to closer stars being mis-assigned large distance values. It would therefore be useful to know how a 0.1 dex

shift (for instance) in the mean metallicity over a large area of the halo would propagate into the deduced density in the [60-80]kpc range.

4) My reading of Figure 2 is that the detection is less than a 2-sigma enhancement above the - already idealized - smooth halo model. Taking into account the possible presence of contamination and/or metallicity variations may lower the actual significance even further. Hence I feel that the detection should be presented as "tentative".

5) My impression from reading the manuscript and from performing a "by-eye" smoothing of the data in Extended Data Figure 3b, is that the signal seen so strikingly in Figure 1a is actually significantly influenced by the adopted Gaussian kernel, and how this kernel deals with the edges of the masked-out map. Is this the case? How are the masked-out regions treated? In what coordinates is the smoothing performed?

6) It should be noted that the trailing arm of the Sagittarius dwarf galaxy runs straight through the middle the claimed southern wake. Admittedly, this structure is not as distant as it is in the northern hemisphere, but being close, it contaminates a wider band on the sky. Also, it has a different metallicity distribution to the average halo which, together with the large errors on the distance estimates, may be problematic given points (2) and (3).

7) For reproducibility, the Solar reflex correction should be spelled out or a reference with the exact procedure should be given.

Author Rebuttals to Initial Comments:

We thank the referees for their very thoughtful and helpful comments. All changes to the manuscript have been highlighted in red. In addition to a point-by-point response to their specific comments below, we have made the following significant changes to the manuscript.

1. We have updated ED Fig 1 to reflect more accurately the main sample, i.e., we changed to 12<W1<15 and pi<0.05 mas. We also made some minor tweaks to the color selection box to avoid possible contamination at the blue end as per the suggestion from the first referee.

2. ED Fig 2 has been updated to include the parallax cut in the main sample. This was inadvertently omitted, which resulted in some foreground contamination. This contamination did not affect the mean distances but did result in large scatter in the photometric distances for some objects. The scatter is now much smaller, providing additional evidence that there is little contamination in our sample.

3. We have added a new figure (now ED Fig. 4), that includes an annotated version of the all-sky map.

4. Fig 2 has been updated in several respects - first, the region definitions have changed in order to “capture more of the signal”; the error bars are now smaller and more faithfully reflect the intuition in Fig 1 (namely that the wake signal is quite strong). We have also de-emphasized discussion of the East-West Northern asymmetry, as that signal is <2 sigma, and instead focus on the more robust feature that is the global Northern overdensity.

We hope that with these changes the article is now acceptable for publication.

Sincerely,
Charlie Conroy, on behalf of the authors.

Referee #1:

- The authors will agree with me that the all-sky response should also be expected in other halo tracers beyond M-giants. Catalogs of K-giants, BHBs and RR Lyrae are publicly available, and I would strongly recommend the authors to analyze them in a similar fashion. This paper would become _much_ stronger if the authors find a similar feature in the distribution of K-giants, BHBs and RR Lyrae. Conversely, the authors may find that the signal only arises in the M-giant population, which would indicate that the anisotropic distribution of stars is not connected to the LMC, or that some flaw exists in the analysis. In my opinion, such a comparison would provide a crucial test of their results.

REPLY:

We agree with the referee's remark that it would be very helpful to see the signal in other samples. A few comments are in order here:

1. This was not sufficiently clear in the paper, but our sample is in fact composed of K giants, not M giants. Specifically, for our adopted metallicity of [Fe/H]=-1.5, the color range
1.3<BP-RP<1.8
corresponds to an RGB temperature range of 3800<Teff<4500 K -- squarely in the K giant regime. So this sample is indeed a sample of K giants. The Methods section has been updated to clarify this point.

2. Unfortunately, there does not exist all-sky catalogs of any other halo tracer at the depths (distances) required for this work. As the 2nd referee notes, the key breakthrough in this paper is the photometric parallax and giant selection technique based on all-sky photometric datasets. The deep BHB and RR Lyrae catalogs that are available are derived either from SDSS or PS1 imaging, and as such are restricted to Dec>-25. They thus do not provide the necessary
sky-coverage to verify a) the collective response, and b) the connection between the local wake and the LMC. Previous work by Belokurov et al. (2019), which reported the Pisces Plume and its tentative association with the LMC wake was in fact based on the PS1 RR Lyrae data. As noted in the paper, the correspondence between the location of the Pisces Plume and our reported LMC wake is strong confirmation that the signal is not confined to our sample. The SDSS BHB sample is difficult to work with owing to contamination from blue straggler stars and photometric uncertainties in the u-band, which become significant

at the faint magnitudes required to study the R>60 kpc halo (see Deason et al. 2014, in which a full forward-modeling of the dataset was required to derive the density profile at R>50 kpc).

RR Lyrae catalogs do exist from Gaia DR2 (e.g., Clementini et al. 2019), but those catalogs are incomplete at the faint end (G>19) - precisely the magnitude range needed to study the R>60 kpc halo with such stars. Mateu et al. (2020) show the completeness of the Gaia RRL catalogs and there is substantial large-angle structure across the sky (including near the location of the wake), presumably due to the Gaia scanning pattern (see their Fig 12 - RRL of course require multiple visits to obtain an accurate light curve, so the completeness will vary with the number of Gaia visits). We anticipate that future data releases will enable a cleaner Gaia RRL sample at the depths required to confirm our reported signal.

- Fig. 1 the authors claim good agreement between data and simulation. But I have found quite strong discrepancies:

* the model predicts an over-density at l>0, b>0, but the data shows an under-density.
* the amplitude of the global response is Delta rho/<rho>~0.8, but the model predicts Delta rho/<rho>~0.1.
* the amplitude and location of the wake is Delta rho/<rho>~0.8, but the model predicts Delta rho/<rho>~0.2.

To my surprise, the text literally states that "the simulation reproduces the location and general behavior of the wake to a remarkable degree", which in my opinion is not supported by the results. I am not sure what to take from this. Either the models are not realistic, or the over-densities are not caused (either entirely or partially) by the LMC. Both alternatives weaken the main conclusion of the paper.

REPLY:

One needs to remember that the simulation is a genuine ab initio prediction of the effect of the LMC on the MW halo, and was in no way tuned to reproduce these results. Given the large available parameter space of the LMC mass and orbit and the MW mass, we do find the overall agreement quite striking. The text clearly indicates the multiple shortcomings of these specific simulations and points toward direction for future improvement. Moreover, note that Fig 2 has been updated in several respects as noted above.

We have modified the text in question to make this point more clearly: "The simulation is a genuine prediction and was not calibrated in any way to reproduce the observed features. The overall agreement between the simulation and observations is therefore quite striking.

Unsurprisingly, given the large available parameter space, the agreement is not perfect."

- How do the results depend on the choice of initial conditions for the stellar halo? Is it possible to mimic the anisotropy shown in Fig 1 by choosing a triaxial halo with the right orientation ?

REPLY:

We have added the following text in the new “simulations” section of the methods: “The detailed structure of the halo response to the LMC’s passage has been recently quantified using Basis Function Expansions\cite{Garavito-Camargo20}. The halo response to the LMC has a strong amplitude in odd l modes. In contrast, triaxial, oblate and prolate halos have no odd l terms, only even terms. As such, the deformations to the MW’s halo caused by the LMC cannot be mimicked by a triaxial halo.”

- The analysis relies on a severe distance cut 60<r/kpc<80, but I cannot find any clear justification for this particular choice. Are they using the N-body model to inform them on the distance range to inspect? Are the two over-densities still visible without this interval?

REPLY:

We have expanded the R_gal cut to 60-100 kpc. This results in a slightly larger sample (which was offset by our more conservative color selection), resulting in a nearly unchanged total sample size. Importantly, with all of these changes, the results remain unchanged.

We have also inspected plots at 50<R<60 kpc and the local wake and collective response features are clearly visible there. We decided to keep the sample restricted to R>60 kpc in order to minimize contamination from Sagittarius and other features that reside at smaller R_gal. Future data from Gaia will deliver much more precise proper motions, which may enable more thorough separation of Sagittarius and disk star contamination from these maps, which would enable more robust comparisons at distances <60 kpc.

We have added the 50-60 kpc map at the end of this report.

- The coordinates need some clarification. The text claims that the coordinates are

Galactocentric but not centered at the sun? Is longitude=0 oriented with the Sun? Or is longitude 180? Would it not be more intuitive to use spherical coordinates phi and pi/2-theta instead of re-defining l and b?

REPLY:

We are following the coordinate system described in Garavito-Camargo et al. (2019). Specifically, instead of centering the coordinate system at the Sun's location (-8.2,0.0,0.0) kpc, the system is shifted to (0,0,0). In this system the LMC-induced features in the simulation are easier to interpret. Basically, this is what the sky would look like if the Sun was at the center of the Galaxy. We stress that the observed features are clearly evident in the usual Galactic coordinate system.

- It would be _extremely_ helpful to add longitude and latitude lines to Fig 1 and Ext Material Figs 3, 4 and 5.

REPLY:

We have done this in all figures, as well as labels to the lines in Fig 1.

- if the coordinates are centered at the Galaxy, why is there an over-density in Extended Data Fig 4 at l=+- 180 deg.? if the halo is axisymmetric, I would expect no dependence on longitude.

REPLY:

We thank the referee for catching this. There were several issues with the R18 mock as presented in ED Fig 4, including not correcting for dust extinction (the mock data had SFD dust extinction), and there was an inadvertent selection in d_helio in the parent catalog. The latter resulted in an enhancement in stars toward the Galactic anticenter (l~180) which is closer to the Sun than the l~0 Galaxy. These issues have all been fixed and the plot now looks as-expected.

- The analysis assumes that the stellar halo is made of stars with a metallicity Fe/H=-1.5 and an age of 10 Gyr. However, the stellar halo shows a non-negligible spread in metallicities and ages, so adopting delta-functions may bias the results in a non-trivial way. How sensitive is the location/amplitude of the over-densities to the choice of age/metallicity? how would it change if full distributions (e.g. Gaussians) are used instead of delta functions?

REPLY:

We note that the adopted age+metallicity of the halo is used only for determining the specific relation between color and absolute magnitude in order to estimate photometric distances. As noted in the text, the adopted age has very little effect on the distances for these old ages (7% change for a factor of two change in age). The metallicity of course has a larger effect. We have added text in the Methods to quantify this: "For example, assuming a fixed color of
$BP-RP=1.4$, increasing the metallicity to [Fe/H]$=-1.0$ would result in 25\% closer distances, while assuming [Fe/H]$=-2.0$ would place the distances 14\% further away, compared to our fiducial isochrone"

But we also emphasize that distances are only used to place stars in broad radial bins. So even 20% systematic uncertainties will not have a significant impact on our results. And critically, the test in ED Fig 2 demonstrates no systematic uncertainties at the 20\% level.

- The authors use stars at latitudes |b|>10 degrees, but only show photometry of stars at b>45 deg (Ext.data Fig 1). How does the photometry look for stars with 10<|b|<45 ? Does the purity of the sample increase or decrease at low latitudes?

REPLY:

As expected, closer to the plane the dwarf sequence rises in prominence in this color-color diagram. We include at the end of this report a version of ED Fig 1 for 20<b<25 for illustrative purposes. The two sequences are still clearly visible. The cooler portion of the giant sequence is less visible - this part of the giant sequence is largely populated by Sagittarius stream stars which are not prominent in this particular Galactic latitude window.

- The authors claim that the sample has high purity, but one can see by eye that there is significant contamination at RP-W1<1.8. How much does the result depend on this cut? could the authors show how Fig 1 changes if the cut is placed at RP-W1> 1.9?

REPLY:

We agree that the separation at the bluest colors is likely less clean than at redder colors. Closer inspection of the distribution of the final set of stars at 60-100 kpc reveals that most of them are redder than RP-W1>1.8, so we decided to make the color selection more conservative. In particular, notice in ED Fig 1 (and related text) that the blue limit has been set to RP-W1=1.8. All plots have been updated to reflect this change.

Furthermore, of the new sample (updated color cuts and expanded R_gal range), only 48 stars lie at 1.8<RP-W1<1.9, so even if this range suffers greater contamination than the redder colors, it is of little concern for the results presented here.

- What is the completeness of the sample? Can the author guarantee that the fraction of M-giants detected is homogeneous across the sky? This can hardly be the case if we consider, for example, extinction. If completeness is not spatially homogeneous, how is this taken into account in the maps constructed in Extended data fig 3?

REPLY:
The Gaia DR2 catalog is ~100% complete at 12<G<18, especially when considering that we avoid the Galactic plane (|b|<10). The unWISE photometric catalog is also ~100% complete over the range 12<W1<16, as demonstrated in Schlafly et al. (2019). We have added the following text in the Methods section to clarify this point: "For the WISE data, we restrict to
$12<W1<15$, a magnitude range that is nearly 100\% complete\cite{Schlafly19} and has formal
uncertainties of 0.01 mag at the faint end of this range."

Beyond |b|>20, the mean E(B-V) for our final sample of stars is 0.042. As quoted in the text, the reddening coefficient for W1 is 0.18, so the mean extinction in W1 is 0.0078 - very small indeed. Of course, we correct for this extinction, so any systematic uncertainties due to extinction should be negligible.

- When the authors compute the average counts per pixel, do they take into account that a significant fraction of the sky is masked? For example, can the authors exclude that the apparent overdensity at l>0, b>0 simply arises because most of the quadrant containing the MCs has been masked? How is the average density _exactly_ calculated?

REPLY:

Indeed, smoothing a masked image is always a challenge. One really has only two options: to either remove the pixels that are masked in the smoothing, or to include the masked pixels, e.g., as 0.0 values. The former is not really an option with a smooth kernel, as we use here. So we have adopted the latter. It is for this reason that we were very careful to make an
apples-to-apples comparison between the realistic N-body simulation and the R18 smooth model. Specifically, comparison of Fig 1 right panel to ED Fig 5, and also comparison of the left and right panels of Fig 4, explicitly demonstrates the effect of not only the selection but also the masking+smoothing effect.

The average density is computed in the unmasked region *before* smoothing. So the exact operation used to process all of the maps is:

1. compute counts in each pixel
2. apply mask
3. compute average
4. smooth

We have added the following text to the Methods section clarifying these issues:

"When computing smoothed maps (as in e.g., Fig. 1), we took the following steps. First, bin the stars into pixels. Second, apply the mask in both proper motions and coordinate space. Third, compute the average density and divide the map by that average. Fourth smooth the map by a Gaussian kernel. Smoothing a masked map creates challenges near the masked regions. As we are using a continuous kernel, we opted to simply set the masked regions to 0.0 for the masking procedure. This is the default behavior of the \texttt{healpy} routines used in this work. Because the exact same procedure was applied to both the data and models, the comparison between the two should be robust."

- The number count in Extended data fig 3 ranges from 0 to 3. This means that the statistics governing the density maps (prior to smoothing) is deeply Poissonian. How is this taken into account when constructing maps? Also, is the plot in Extended data fig 3 sampling the halo component with Poissonian (low-number) statistics, or are the authors simply using all the particles in the model? To compare apples with apples, my suggestion would be to use the same number of particles as stars in the catalogue (~720) and smooth data and simulations in the same way (including masks) in order to allow a meaningful comparison of Delta rho between models and data.

REPLY:

The maps are plotted as-is, with no special consideration for Poisson statistics. Part of the purpose of smoothing the maps with the Gaussian kernel is to average out the Poisson fluctuations. This is one reason we provide both the smoothed and unsmoothed maps - to enable comparison between the raw counts and the smoothed, averaged version.

In all comparisons between data and simulations the maps are processed in the exact same way, including the masking and smoothing (in that order). In both the LMC+MW simulation and the smooth model, we sample the profile to match the data, using all of the particle data. We do not see the purpose of introducing noise into the simulations by down-sampling to the observed number of stars in the data - such a procedure will not introduce large-scale behavior in the simulations.

Poisson uncertainties *are* included in Fig 2. It is there that a more meaningful quantitative comparison can be made between the simulations and the data.

-The colour bars in Extended Data Figure 3 must be wrong. There are pixels with 3 stars in the right panel that don't have 3 stars in the left panel. Where did these stars come from?

REPLY:
We thank the referee for their attention to detail! We made an error in the creation of the unsmoothed and unfiltered map (left panel), in which we neglected to re-project the coordinate system to be centered on the Galactic Center. This has been fixed in the updated version, with results as-expected (the color-bar was correct all along).

-In general, the description of what they did with the N-body model is not anywhere near robust enough. What assumptions go into building the LMC-MW models? Are they treating stellar halo as a compilation of individual particles or re-weighting DM particles?

REPLY:

We have added a new section to the Methods titled "Simulation Details" that gives pertinent details of the simulations presented in Garavito-Camargo et al. (2019) and our treatment of the particle data. See also our response to the next comment.

- What are the Initial Conditions for the MW stellar halo? could the authors show the radial profile of M-giants and the stellar halo model? The authors claim that they sample the model to have the same radial distribution as the data, but I am not convinced this will give a fair (or unique, at 720 stars) answer. For example, how could the contrast change by sampling from stellar halo models with different (initial) slopes?

REPLY:

There is no separate stellar halo in the Garavito-Camargo et al. simulations. In that work they employ a re-weighting scheme to approximate a stellar halo (their Section 3.4). We have considered both their re-weighting scheme and the raw DM particles and in both cases have re-sampled the particles to match the observed number density profile. The results are in all cases unchanged, aside from very slight differences in Fig 2, since the strength of the wake signal does depend on distance. Given the relatively crude distances available in the data, we did not pursue this avenue further. We have noted this test in the Simulations section.

- why the distance cut at r>60kpc does not remove all disc stars? Does it mean that the sample contains misclassified disk stars as M-giants?

REPLY:

Yes, some stars, the vast majority of which are at |b|<20 and |ell|<90, evidently have incorrect distances. This is likely due to the incorrect metallicity adopted (as disk stars are metal-rich), and/or variations in the reddening law which makes our reddening corrections inaccurate. The PM cut efficiently removes those stars, but it comes at a cost - the PMs are somewhat noisy, so the PM cut does remove stars from the wake (as you can see by comparing the left and right panels of ED Fig 3). We felt it was safer to have a pure sample, than a noisy and complete one. Future data from Gaia, with more precise PMs, will be helpful on this front.

- I am not sure how the proper motion cuts are chosen to remove disk stars. Could the authors show how it works?

REPLY:

The selection was done on a trial and error basis by studying the observed, reflex-corrected proper motion space. There is, as always, a trade-off between completeness and purity. We can report that, of the stars removed because of the proper motion cut,

only 13% reside at
|b|>30 deg, where most of our signal resides. One can see this by direct inspection of the left and right panels of ED Fig 3. The proper motion selection is the only one that reduces the number of stars in regions of the map not explicitly masked in grey. Future data releases from Gaia will enable more stringent selections via higher SNR PMs.

Referee #2:

This is a very interesting and well-written study, reporting the discovery of the long sought-for response of the Galactic halo due to the arrival of the Magellanic system. Showing the reality of gravitational wakes is of extreme importance for our understanding of the dark matter - indeed it would provide smoking-gun evidence for the particulate nature of this unseen component of our Universe. I am aware of many attempts to try to find the LMC wake (including by my own team!) that did not result in a positive detection. The key insight that is brought to bear on this problem in the present manuscript is shown on Figure 1 in the Extended Data, which illustrates the new photometric parallax technique they have found (and which I have verified). In my opinion, this is an important result in itself.

On the other hand, I am concerned about the reality and statistical significance of the claimed detection of the LMC/SMC wake. I would like to see the following additional analysis and clarifications before recommending publication:

1) While I agree that selecting in the distance interval of 60 to 80 kpc seems like a reasonable choice, what do adjacent distance cuts reveal? From my own selections trying to reproduce the present result, I see that there are many more stars in the 50-60kpc range.

REPLY:

Motivated by both referee's comments on this topic, we have extended our fiducial R_gal range to 60-100 kpc. This results in only a modest increase in the sample size (partly counterbalanced with the more restrictive color selection).

We have also inspected plots at 50<R<60 kpc and the local wake and collective response features are clearly visible there. We decided to keep the sample restricted to R>60 kpc in order to minimize contamination from Sagittarius and other features that reside at smaller R_gal. See the end of this report for a map at 50-60 kpc.

----

2) I applied the author's selection (but without the complex spatial window) to the public LAMOST DR5 sample. Interestingly, there is a substantial metallicity range of these potential "halo" stars: -1.7 < [Fe/H] < -0.4, I presume due to the presence of sub-structure or contamination. This suggests to me that contamination may be a concern.

REPLY:

We have verified with H3 Survey data that implementing the color and parallax cuts results in a metallicity range of about -1.5<[Fe/H]<-0.5 with a std dev. of 0.25. So there is indeed some metallicity spread, as the referee notes. However, we refer back to ED Fig 2, which shows good photometric distances for GCs and dwarf galaxies spanning a wide range in [Fe/H] (now noted in the caption). So for these reasons we do not see the metallicity range as a concerning source of contamination.

----

3) The smooth halo model that is used as a comparison presumably assumes that the stellar populations in the halo are also perfectly mixed. However, the large number of known halo substructures demonstrates that this assumption is not perfectly valid, and many chemically-different halo features may yet lie just below present detection limits. Given the halo's steep density profile, one may then worry that small deviations from this assumption could lead to closer stars being mis-assigned large distance values. It would therefore be useful to know how a 0.1 dex shift (for instance) in the mean metallicity over a large area of the halo would propagate into the deduced density in the [60-80]kpc range.

REPLY:

We have added text in the Methods section describing the effect of a much larger metallicity change (+/-0.5 dex), which amounts to a +25/-15% change in distance. Recall that distances are only used to sort stars in quite broad radial bins, so we do not expect any potential errors in distance to play a role in the analysis. Moreover, in answer to an earlier comment, the signal clearly does not depend on the exact radial range of interest - the signal is present at 50-60, 60-80, and 80-100 kpc - which further suggests to us that any possible distance systematic error is unlikely to impact the results.

----

1) My reading of Figure 2 is that the detection is less than a 2-sigma enhancement above

the - already idealized - smooth halo model. Taking into account the possible presence of contamination and/or metallicity variations may lower the actual significance even further. Hence I feel that the detection should be presented as "tentative".

REPLY:

We have made several improvements to Fig 2 in part to address this concern. We have enlarged the regions used to define the ratios, which has allowed us to “capture” more of the signal. The density contrast in the wake is now a ~6 sigma detection. This also conforms better with our intuition from Fig 1 and ED Fig 3, where the signal is very obvious by-eye.

----

2) My impression from reading the manuscript and from performing a "by-eye" smoothing of the data in Extended Data Figure 3b, is that the signal seen so strikingly in Figure 1a is actually significantly influenced by the adopted Gaussian kernel, and how this kernel deals with the edges of the masked-out map. Is this the case? How are the masked-out regions treated? In what coordinates is the smoothing performed?

REPLY:

Please also see our reply to the 1st referee who raised similar points. The smoothing was performed in the Galactocentric coordinate frame shown in the maps. The masked regions are set to 0.0 so edge-effects will appear in the maps. This is one reason why we made sure to apply the exact same procedures to the models, to enable direct comparison. At the end of this reply we have included maps smoothed by 10 and 20 deg. (FWHM) for comparison purposes. We chose 30 deg. as that was the smallest scale that emphasized the smooth behavior of the halo.

----

1) It should be noted that the trailing arm of the Sagittarius dwarf galaxy runs straight through the middle of the claimed southern wake. Admittedly, this structure is not as distant as it is in the northern hemisphere, but being close, it contaminates a wider band on the sky. Also, it has a different metallicity distribution to the average halo which, together with the large errors on the distance estimates, may be problematic given points (2) and (3).

REPLY:

We have added the following text to the Methods section acknowledging this issue: "The plane of Sagittarius extends to $b<0^\circ$ at $l\approx0^\circ$, which intersects perpendicular to the local wake. The Southern arm of Sagittarius is at smaller $R_{\rm gal}$, so we do not need to mask this feature in our maps."

We are confident that we have assigned the correct distance to the Sgr arms because we see them at smaller Rgal. Moreover, in ED Fig 2, there are galaxies and GCs ranging from [Fe/H]=-0.5 to -2.2 and all have good recovered distances. We have added a comment to this effect in the caption to that figure.

----

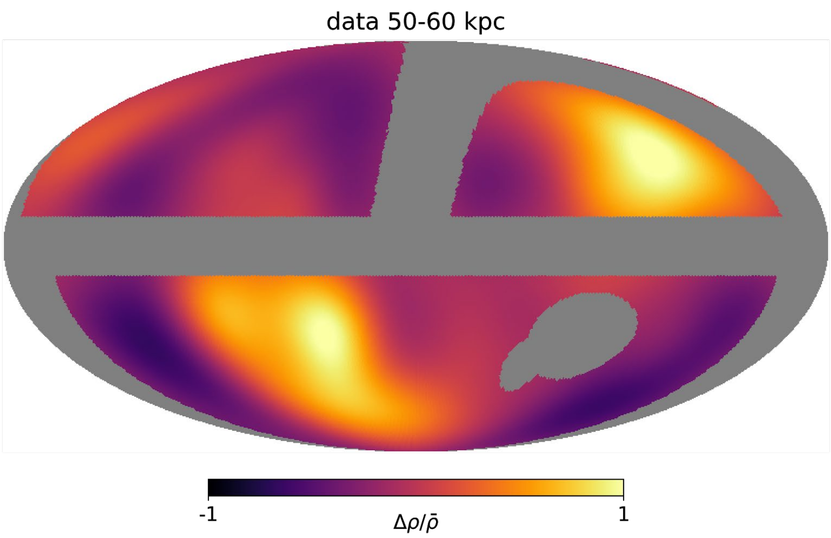
2) For reproducibility, the Solar reflex correction should be spelled out or a reference with the exact procedure should be given.

REPLY:

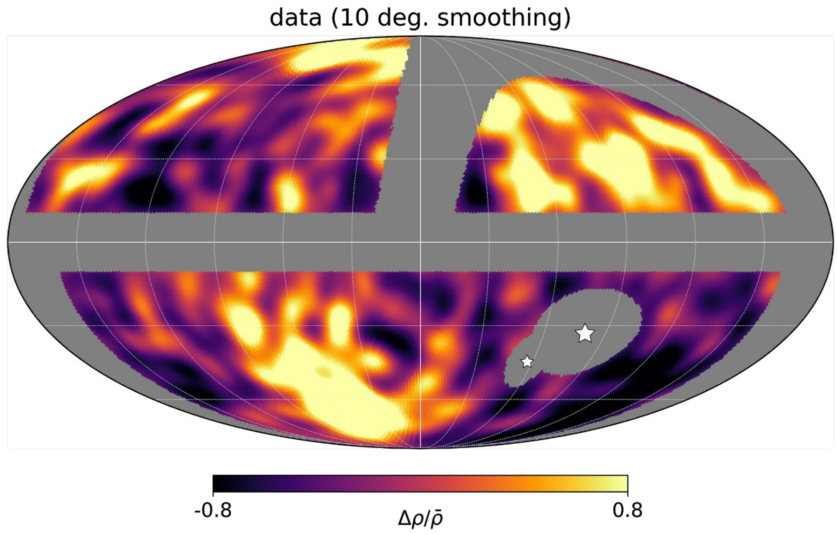
We have added the following text as a footnote in the Methods section detailing the solar reflex correction:

We adopt the Galactocentric frame implemented in \texttt{Astropy v4.0} \citep{astropy1, astropy2} which has the following parameters: $R_{0}=8.122$ kpc \citep{Gravity19}, $[V_{\rm{R,\odot}}, V_{\rm{\phi,\odot}}, V_{\rm{Z,\odot}}]=[-12.9, 245.6, 7.78]$ km s^{-1} \citep{Drimmel18}, $Z_{\rm{\odot}}=20.8$ pc \citep{BennettBovy19}. We use the \citep{gala} package's \texttt{reflex_correct} by setting radail velocities to zero to account for the imprint of the solar motion on our proper motions.

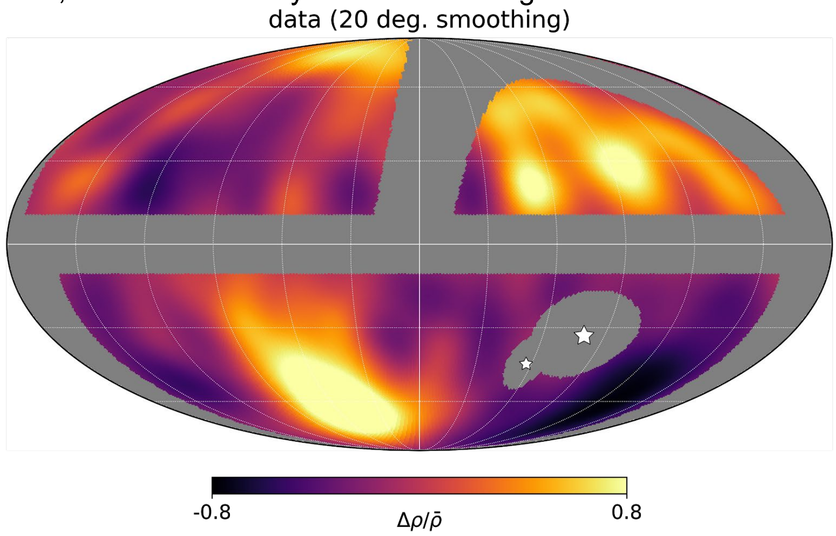
Supplementary Figures



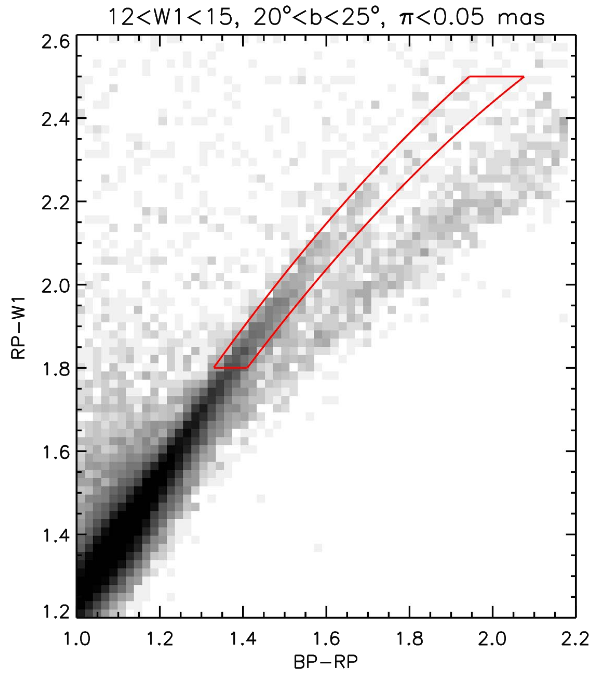
Map of the data for stars with 50<R_gal<60 kpc. Note that the wake is clearly visible in this distance slice. That the wake and northern over-densities are clearly visible in two independent portions of the data adds further evidence to the reality of these features.



As in Fig 1 left panel, now smoothed by a FWHM=10 degree kernel.



As in Fig 1, now smoothed by a FWHM=20 degree kernel.



As in ED Fig 1, now for 20<b<25 deg. Note that the giant sequence is clearly visible, though the dwarf sequence has increased in prominence as we observe closer to the plane.

Reviewer Reports on the First Revision:

Referee #1 (Remarks to the Author):

I thank the authors for addressing some of the points raised in my report. Unfortunately, some important questions remain unanswered.

GENERAL COMMENTS ----

Comparison against the N-body simulation continues to show important discrepancies with the data. In particular, the theoretical model predicts a clear overdensity at l>0, b>0, whereas the data show an underdensity. The authors claim that one should not pay too much attention to this mismatch given that (i) the model has not been tailored to this particular data set, and (ii) the similarities outweight the discrepancies. I find this argument not particularly compelling, given that one could equally argue that the clear mismatch invalidates _this_ particular model, and thus any conclusion drawn from it.

Taking this issue at face value means that the detection is tentative at best, as pointed out by the second referee. In my opinion, the 6-sigma detection claimed by the authors simply shows that the K-giant distribution is not isotropic, but it does not establish a physical connection with the LMC necessarily. Or to put it in a different way: correlation with a single simulation does not mean causation. To prove that connection, the authors need to identify a genuine prediction of the model that can be validated with independent data. The authors claim in several places that more precise proper motions will help to make the signal stronger, but the issue here is not about measuring the anisotropy with higher accuracy, but determining whether or not the anisotropy in the K-giant distribution is caused by the LMC.

In this regard, the easiest way to validate this model is by comparing different families of halo tracers, and I strongly urge the authors to at least try (see below). The authors claim that full sky coverage is essential for this task. Yet, the new Extended Data Figure 4 shows that SDSS and Pan-STARRS cover much of the area of interest. Even if the LMC quadrant is missing, I do believe that the remainder 3 quadrants contain sufficient information to determine whether or not BHBs and RRL stars show a similar degree of anisotropy as the K-giants. For example, the clear south/north asymmetry at l>0 should be clearly visible in SDSS and Pan-STARRS.

SPECIFIC COMMENTS ----

**Authors :
As the 2nd referee notes, the key breakthrough in this paper is the photometric parallax and giant selection technique based on all-sky photometric datasets.
The deep BHB and RR Lyrae catalogs that are available are derived either from SDSS or PS1 imaging, and as such are restricted to Dec>-25. They thus do not provide the necessary sky-coverage to verify a) the collective response, and b) the connection between the local wake and the LMC.

Referee: I disagree. Most of the collective response signal is located in the 3 quadrants covered by SDSS and Pan-STARRS. Given that there is no clear physical link between the overdensities and the LMC, I would argue that the 4th quadrant may not be essential to test whether or not the collective response is visible in other tracer families.

Public catalogs of BHBs and RRL stars covering those 3 quadrant are available. I strongly suggest the authors to use them.

**Authors:
The SDSS BHB sample is difficult to work with owing to contamination from blue straggler stars and photometric uncertainties in the u-band, which become significant at the faint magnitudes required to study the R>60 kpc halo

Referee: Significant progress has been made to identify and remove BS contaminants, see Lancaster et al. (2019)

**Authors:
RR Lyrae catalogs do exist from Gaia DR2 (e.g., Clementini et al. 2019), but those catalogs are incomplete at the faint end (G>19) - precisely the magnitude range needed to study the R>60 kpc halo with such stars.

Referee: Rimoldini et al. (2919) include RRL that reach out to 120 kpc. Although it is true that the sample is incomplete at G>20, this should not be an issue to determine whether or not the distribution is anisotropic on the sky given that the authors are not interested in the radial profile and only introduce a broad distance cut at 60 kpc.

**Authors:
Mateu et al. (2020) show the completeness of the Gaia RRL catalogs and there is substantial large-angle structure across the sky (including near the location of the wake), presumably due to the Gaia scanning pattern (see their Fig 12 - RRL of course require multiple visits to obtain an accurate light curve, so the completeness will vary with the number of Gaia visits).

Referee: the fluctuations introduced by the Gaia scanning law have significantly smaller amplitudes than the Poisson fluctuations in the number counts shown in Extended Data Figure 3. I believe that a 30 degree Gaussian kernel will be able to remove those artefacts quite efficiently.

In addition, the authors may want to analyze the K-giant catalogue of Xue et al (2014) built from SDSS/SEGUE data. The comparison against WISE catalog could be very interesting for several reasons, e.g. (i) in order to validate the results shown here, and (ii) to quantify the importance of having access to the LMC quadrant.

I can see no strong reason for not using those data sets, which are all public. Could the authors produce maps of (SDSS) K-Giants, RRLs and BHBs using a similar analysis as presented here? This could be a key test of the scenario proposed here.

**Authors:
We have added the following text in the new "simulations" section of the methods: "The detailed structure of the halo response to the LMC's passage has been recently quantified using Basis Function Expansions\cite{Garavito- Camargo20}. The halo response to the LMC has a strong amplitude in odd l modes. In contrast, triaxial, oblate and prolate halos have no odd l terms, only even terms. As such, the deformations to the MW's halo caused by the LMC cannot be mimicked by a triaxial halo."

Referee: The authors are missing my point. The issue is not whether the perturbed halo in the N-body models of Garavito-Camargo et al (2020) requires odd-l modes. My question is whether the _observed_ anisotropy of the K-giant distribution can be mimicked by a aspherical halo whose symmetry axis if off the disk plane. For example, the figure attached here shows a prolate distribution of stars inclined 30 degrees from the disk. By eye, I would argue that this model may fit *better* the observed anisotropy than the halo model provided by the authors

[Figure Redacted]

Can the authors rule out this scenario with the data at hand?
They are welcome to apply a spherical harmonics decomposition to the K-giant data sets if they believe that this helps to answer my question.

[why the distance cut at r>60kpc does not remove all disc stars? I am not sure how the proper motion cuts are chosen to remove disk stars. Could the authors show how it works?]

Authors:
The selection was done on a trial and error basis by studying the observed, reflex-corrected proper motion space. There is, as always, a trade-off between completeness and purity. We can report that, of the stars removed because of the proper motion cut, only 13% reside at |b|>30 deg, where most of our signal resides. One can see this by direct inspection of the left and right panels of ED Fig
3. The proper motion selection is the only one that reduces the number of stars in regions of the map not explicitly masked in grey. Future data releases from Gaia will enable more stringent selections via higher SNR PMs.

Referee: I am still confused about this approach. The text claims that the proper motion cut reduces the catalogue size from 3,878 stars down to 1,293 (this is after introducing a distance cut and masking the disk plane). This means that 2/3 of the WISE sample have biased distances and are potential contaminants, which makes the cut in proper motion space a rather crucial step. The authors claim that the proper motion cuts are made by eye on a trial and error basis. I am not convinced that this is the best way to identify disk stars, nor am I sure to what extent the cut is affecting the results.

-Could the authors show how the signal varies for different proper motion cuts? - Can those cuts be justified on dynamical grounds (i.e. by comparing them against disk orbits?)
-Would it be more sensible to apply cuts in velocities, rather than proper motions? (here, I presume that nearby disk stars misclassified as K-giants have unreasonably large velocities).

Referee #2 (Remarks to the Author):

Please find below my report on the revised version of "Detection of the All-Sky Response of the Galactic Halo to the Magellanic Clouds", by C. Conroy and collaborators.

I find the new version of the manuscript to be significantly improved, although I still suspect that the authors are over-selling the result. Nevertheless, the paper is interesting and will undoubtedly stimulate discussion in the community. I find it very reassuring that the structure remains visible in the 50-60kpc distance interval as the authors show in their reply.

Most readers will probably build a mental picture of the Milky Way's response to the Magellanic system from Figure 1, and its comparison to ED Figure 5. Yet clearly ED Figure 5 was smoothed, and then masked, not masked and then smoothed, as the authors claim!
The revised manuscript states that the authors apply the mask, divide by the average, and then smooth. Why then do they not assign a value of unity to the masked region, to avoid excessive artifacts? The presently adopted procedure will have the effect of artificially enhancing blobs ~30 degrees from the masks. Because of this, I feel that the density maps, and Figure 1 in particular, remain misleading to the community.

In my previous report I asked the authors in which coordinate system the smoothing was performed. Their response did not really answer the concern behind my question: was the smoothing done properly (i.e. accounting for great-circle distances)?

Also in my previous report, I asked about the effect of metallicity variations in the halo. This has been answered to some extent, and the authors state the distance changes corresponding to different metallicity assumptions. I appreciate that precise distances are not needed to place stars into wide bins. However, there is a strong density profile present in the halo, so it is not inconceivable that spatially-dependent metallicity variations could create apparent spatial overdensities. It would be useful to readers for this possibility to be quantified.

Having tried to reproduce the authors' sample from the public databases, I feel that it is not entirely trivial to recover their dataset. I would therefore strongly recommend that the authors include a table with the 793 stars they have identified. This would an additional useful contribution to the community wishing to follow-up their detection.

Author Rebuttals to First Revision:

We thank both referees for their additional very helpful comments. In particular, we thank the first referee for urging us to explore ancillary datasets. In response to their comments we have added a new section and figure discussing a distant halo map based on RR Lyrae (ED Fig. 7), a new figure demonstrating our proper motion selection technique (ED Fig. 4), and a new table containing the final K giant sample (ED Table 1). In light of the significant changes throughout the manuscript, we felt it was impractical to highlight particular changes.

We have also taken the opportunity to make use of the new Gaia EDR3 data. The much more precise proper motions results in a significantly improved selection of stars in proper motion space (as shown in ED Fig 4). In particular, we are now able to remove Sagittarius stream stars via PM cuts in most cases, thereby significantly reducing the fraction of sky that needs to be masked. The sample has also increased by almost a factor of two (to 1301 stars). As a result of these changes (as well as the adoption of a new colorbar), the key plot in Fig 1 looks slightly different, but the key qualitative results (and quantitative result shown in Fig 2) remain unchanged.

Finally, we have changed the coordinate system from Galactocentric (centered on the Galaxy) to Galactic (the more conventional heliocentric system) because we realized that the removal of highly reddened regions of the sky should be included in our mask (it is now included as such in all plots), and this mask does not translate well into the Galactocentric frame (regions that are near to the plane in Galactic coordinates are projected in some cases to high latitude in Galactocentric coordinates). The text and all figures have been updated to reflect this change.

We hope that with these changes the article is now acceptable for publication.

Sincerely,
Charlie Conroy, on behalf of the authors.

Referee #1 (Remarks to the Author):

GENERAL COMMENTS ----

Comparison against the N-body simulation continues to show important discrepancies with the data. In particular, the theoretical model predicts a clear overdensity at l>0, b>0, whereas the data show an underdensity. The authors claim that one should not pay too much attention to this mismatch given that (i) the model has not been tailored to this particular data set, and (ii) the similarities outweigh the discrepancies. I find this argument not particularly compelling, given that one could equally argue that the clear mismatch invalidates _this_ particular model, and thus any conclusion drawn from it.

Taking this issue at face value means that the detection is tentative at best, as pointed out by the second referee. In my opinion, the 6-sigma detection claimed by the authors simply shows that the K-giant distribution is not isotropic, but it does not establish a physical connection with the LMC necessarily. Or to put it in a different way: correlation with a single simulation does not mean causation. To prove that connection, the authors need to identify a genuine prediction of the model that can be validated with independent data. The authors claim in several places that more precise proper motions will help to make the signal stronger, but the issue here is not about measuring the anisotropy with higher accuracy, but determining whether or not the anisotropy in the K-giant distribution is caused by the LMC.

In this regard, the easiest way to validate this model is by comparing different families of halo tracers, and I strongly urge the authors to at least try (see below). The authors claim that full sky coverage is essential for this task. Yet, the new Extended Data Figure 4 shows that SDSS and Pan-STARRS cover much of the area of interest. Even if the LMC quadrant is missing, I do believe that the remaining 3 quadrants contain sufficient information to determine whether or not BHBs and RRL stars show a similar degree of anisotropy as the K-giants. For example, the clear south/north asymmetry at l>0 should be clearly visible in SDSS and Pan-STARRS.

SPECIFIC COMMENTS ----

**Authors :

As the 2nd referee notes, the key breakthrough in this paper is the photometric parallax and giant selection technique based on all-sky photometric datasets. The deep BHB and RR Lyrae catalogs that are available are derived either from SDSS or PS1 imaging, and as such are restricted to Dec>-25. They thus do not provide the necessary sky-coverage to verify a) the collective response, and b) the connection between the local wake and the LMC.

Referee: I disagree. Most of the collective response signal is located in the 3 quadrants covered by SDSS and Pan-STARRS. Given that there is no clear physical link between the overdensities and the LMC, I would argue that the 4th quadrant may not be essential to test whether or not the collective response is visible in other tracer families.

Public catalogs of BHBs and RRL stars covering those 3 quadrant are available. I strongly suggest the authors use them.

**Authors:

The SDSS BHB sample is difficult to work with owing to contamination from blue straggler stars and photometric uncertainties in the u-band, which become significant at the faint magnitudes required to study the R>60 kpc halo

Referee: Significant progress has been made to identify and remove BS contaminants, see Lancaster et al. (2019)

**Authors:

RR Lyrae catalogs do exist from Gaia DR2 (e.g., Clementini et al. 2019), but those catalogs are incomplete at the faint end (G>19) - precisely the magnitude range needed to study the R>60 kpc halo with such stars.

Referee: Rimoldini et al. (2919) include RRL that reach out to 120 kpc. Although it is true that the sample is incomplete at G>20, this should not be an issue to determine whether or not the distribution is anisotropic on the sky given that the authors are not interested in the radial profile and only introduce a broad distance cut at 60 kpc.

**Authors:

Mateu et al. (2020) show the completeness of the Gaia RRL catalogs and there is substantial large-angle structure across the sky (including near the location of the wake), presumably due to the Gaia scanning pattern (see their Fig 12 - RRL of course require multiple visits to obtain an accurate light curve, so the completeness will vary with the number of Gaia visits).

Referee: the fluctuations introduced by the Gaia scanning law have significantly smaller amplitudes than the Poisson fluctuations in the number counts shown in Extended Data Figure 3.

I believe that a 30 degree Gaussian kernel will be able to remove those artefacts quite efficiently.

In addition, the authors may want to analyze the K-giant catalogue of Xue et al (2014) built from SDSS/SEGUE data. The comparison against WISE catalog could be very interesting for several reasons, e.g. (i) in order to validate the results shown here, and (ii) to quantify the importance of having access to the LMC quadrant.

I can see no strong reason for not using those data sets, which are all public. Could the authors produce maps of (SDSS) K-Giants, RRLs and BHBs using a similar analysis as presented here? This could be a key test of the scenario proposed here.

REPLY:

We thank the referee for urging us to consider ancillary datasets. We have carefully considered RR Lyrae as an independent check on our results. We have added a new Section to the Methods, and a new Ex. Data Fig. 7 that shows both the RRL sky map (based on PS1 data from Sesar et al. 2017) and a quantitative comparison to both the K giants and the simulations. We believe that this comparison is valuable and serves to strengthen the conclusions of the paper.

We also explored in some detail the Gaia-based RRL sample, as suggested by the referee. In order to test the viability of the Gaia sample, we compared it to the PS1 sample in the regions where they overlap. The result is shown in Supp. Fig. 2 at the end of this report. The conclusion is that the local wake, so prominent in the PS1 data, is not visible in the Gaia data. We believe that the reason for this is due to the incompleteness in the Gaia RRL catalog.
Namely, the Gaia scanning law imparts substantial, large-angle incompleteness across the sky, at precisely the location of the local wake (as demonstrated in Mateu et al. 2020). Note that RRL at >60 kpc have G~20, which is pushing the faint end of the Gaia RRL catalog. Whatever the reason, it is clear that the current Gaia RRL catalog is not usable for this exercise.

We're not quite sure we understand the value of the SEGUE K giant sample from Xue et al. That sample has 145 stars at R_gal>60 kpc, a factor of 10 smaller than our sample, and with a complex on-sky selection function. We therefore did not pursue this sample further here.

We inspected the cleaned catalog of SDSS BHB stars from Lancaster et al., as suggested by the referee. That sample contains only 6 stars at >60 kpc. Like the RRL, those stars are faint (G>19.5) and so have low SNR proper motions. Owing to the limited depth (in terms of R_gal) and given that the RRL analysis largely confirms the K giant results, we did not pursue the BHB suggestion further.

Again, we thank the referee for these suggestions. We believe that the addition of the RRL sample strongly corroborates the conclusions reached by the K giant sample and retires concerns that the K giant sample is in some way compromised or that the signal is restricted to just one population.

**Authors:

We have added the following text in the new “simulations” section of the methods: “The detailed structure of the halo response to the LMC’s passage has been recently quantified using Basis Function Expansions\cite{Garavito- Camargo20}. The halo response to the LMC has a strong amplitude in odd l modes. In contrast, triaxial, oblate and prolate halos have no odd l terms, only even terms. As such, the deformations to the MW’s halo caused by the LMC cannot be mimicked by a triaxial halo.”

Referee: The authors are missing my point. The issue is not whether the perturbed halo in the N-body models of Garavito-Camargo et al (2020) requires odd-l modes. My question is whether the _observed_ anisotropy of the K-giant distribution can be mimicked by a aspherical halo whose symmetry axis is off the disk plane. For example, the figure attached here shows a prolate distribution of stars inclined 30 degrees from the disk. By eye, I would argue that this model may fit *better* the observed anisotropy than the halo model provided by the authors

Can the authors rule out this scenario with the data at hand?

They are welcome to apply a spherical harmonics decomposition to the K-giant data sets if they believe that this helps to answer my question.

REPLY:

We thank the referee for this suggestion. We include at the end of this report a figure (Supp. Fig. 1) that is a by eye fit to the data using a tilted prolate halo model (this model is inclined by 60 degrees). In a very broad sense, this model captures some of the main elements in the data. However, this simple model fails in at least two important respects: 1) in the south, the shape of the over-density is qualitatively different from the observed local wake. In particular, the overdensity traces a line of constant latitude, while in both the data and the simulations the wake is curved and traces the predicted orbit of the LMC (this is especially evident at b<-75).
And 2) in the North, one would expect to see symmetric overdensities, but this is not what we observe. The Northern overdensity is much larger on the sky, and does not have a shape mirrored by the South. For these reasons we disfavor the tilted halo solution.

[why the distance cut at r>60kpc does not remove all disc stars? I am not sure how the proper motion cuts are chosen to remove disk stars. Could the authors show how it works?]

Authors:

The selection was done on a trial and error basis by studying the observed, reflex-corrected proper motion space. There is, as always, a trade-off between completeness and purity. We can report that, of the stars removed because of the proper motion cut, only 13% reside at |b|>30 deg, where most of our signal resides. One can see this by direct inspection of the left and right panels of ED Fig 3. The proper motion selection is the only one that reduces the number of stars in regions of the map not explicitly masked in grey. Future data releases from Gaia will enable more stringent selections via higher SNR PMs.

Referee: I am still confused about this approach. The text claims that the proper motion cut reduces the catalogue size from 3,878 stars down to 1,293 (this is after introducing a distance cut and masking the disk plane). This means that 2/3 of the WISE sample have biased distances and are potential contaminants, which makes the cut in proper motion space a rather crucial step. The authors claim that the proper motion cuts are made by eye on a trial and error basis. I am not convinced that this is the best way to identify disk stars, nor am I sure to what extent the cut is affecting the results.

-Could the authors show how the signal varies for different proper motion cuts? - Can those cuts be justified on dynamical grounds (i.e. by comparing them against disk orbits?)

-Would it be more sensible to apply cuts in velocities, rather than proper motions? (here, I presume that nearby disk stars misclassified as K-giants have unreasonably large velocities).

REPLY:

In light of the new EDR3 data, we have re-visited our selection in proper motion space in order to remove as much of Sgr as possible in PM space and not in sky-coordinates. Our new procedure is shown in a new Ex. Data. Fig 4 as well as an updated Methods section. With the new PM data and updated selections the fraction of stars removed in the final “zero PM” selection is about 50% (after removing the LMC+SMC and Sgr dSph). Of these remaining stars, nearly all are either associated with a Southern Sgr arm (at l~180) or with the disk at
|b|<20. It is not surprising that we would have some contamination from the disk. Such stars are much more metal-rich and so our photometric distances, which are calibrated at low metallicities, will be more prone to errors. Furthermore, since the disk is ~100x the mass of the stellar halo, even a tiny fraction of contaminants from the disk will result in a non-trivial number of stars in our K giant sample. Thankfully the high SNR proper motions from EDR3 offer an effective filter to remove such stars.

Referee #2 (Remarks to the Author):

I find the new version of the manuscript to be significantly improved, although I still suspect that the authors are over-selling the result. Nevertheless, the paper is interesting and will undoubtedly stimulate discussion in the community. I find it very reassuring that the structure remains visible in the 50-60kpc distance interval as the authors show in their reply.

Most readers will probably build a mental picture of the Milky Way's response to the Magellanic system from Figure 1, and its comparison to ED Figure 5. Yet clearly ED Figure 5 was smoothed, and then masked, not masked and then smoothed, as the authors claim!

REPLY:

The Figures are all processed in the same way, exactly as described in the text. Namely, the maps are masked *and then smoothed*.

The revised manuscript states that the authors apply the mask, divide by the average, and then smooth. Why then do they not assign a value of unity to the masked region, to avoid excessive artifacts? The presently adopted procedure will have the effect of artificially enhancing blobs
~30 degrees from the masks. Because of this, I feel that the density maps, and Figure 1 in particular, remain misleading to the community.

REPLY:

We compute the density contrast where the average density has a value of 0.0, so setting the masked regions to 0.0 avoids artefacts as much as possible. The text was unclear on this point; we have updated the text to clarify that we compute the density contrast (with mean density of 0.0) before smoothing.

In my previous report I asked the authors in which coordinate system the smoothing was performed. Their response did not really answer the concern behind my question: was the smoothing done properly (i.e. accounting for great-circle distances)?

REPLY:

Yes, the smoothing was done properly. The Healpy package was used to make maps and smooth. Healpy relies on HealPix, which divides the sky up into equal-area pixels. For smoothing, healpy first transforms the map into spherical harmonics, smooths in that space, and then transforms back to coordinate space.

Also in my previous report, I asked about the effect of metallicity variations in the halo. This has been answered to some extent, and the authors state the distance changes corresponding to different metallicity assumptions. I appreciate that precise distances are not needed to place stars into wide bins. However, there is a strong density profile present in the halo, so it is not inconceivable that spatially-dependent metallicity variations could create apparent spatial overdensities. It would be useful to readers for this possibility to be quantified.

REPLY:

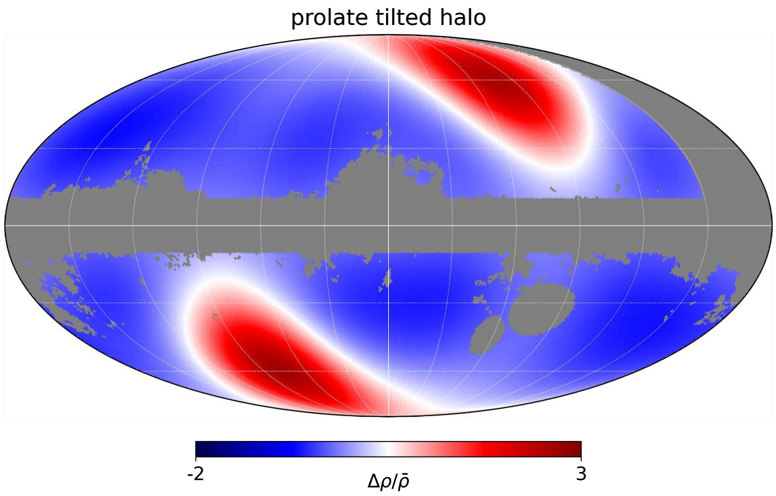
We believe that the issue raised by the referee is unlikely to be an issue in our case for the following reasons. The most metal-rich populations in the halo are [Fe/H]=-0.8 to -1.0, so the error in the photometric metallicities cannot be much greater than 25% (as noted in the Methods). At 60 kpc, this would imply a true distance of 48 kpc, while at 100 kpc this would imply a distance of 80 kpc. But we have already demonstrated in our previous report that the signal remains at 50-60 kpc. We have also split the sample further into 60-80 and 80-100 kpc and the signal remains in those two separate bins. Distance errors would therefore have to exceed 2x in order for a trace metal-rich, spatially asymmetric population to contaminate all of these bins. (The fact that we see a similar signal in the RRL also argues against a metallicity-dependent bias effect). Since we have argued that the photometric distances are more precise than this, we believe that the scenario suggested by the referee is very unlikely to cause the observed behavior.

Having tried to reproduce the authors' sample from the public databases, I feel that it is not entirely trivial to recover their dataset. I would therefore strongly recommend that the authors include a table with the 793 stars they have identified. This would be an additional useful contribution to the community wishing to follow-up their detection.

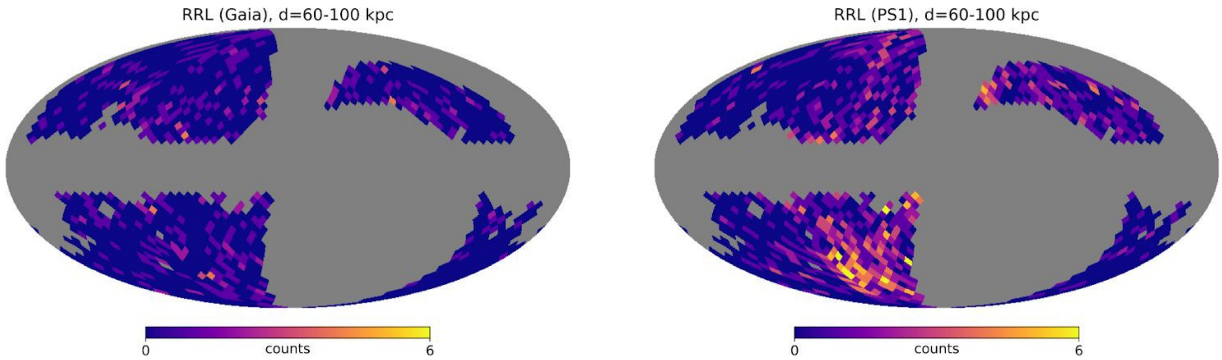
REPLY:

This is an excellent suggestion. We have added a table to the supplementary material.

Supplementary Figures



Supp. Fig. 1. *A simple tilted prolate halo model with parameters chosen to match as closely as possible the data.*



Supp. Fig. 2. *Comparison of RR Lyrae maps at 60-100 kpc based on Gaia data (left panel; Rimoldini et al. 2019), and PS1 data (right panel; Sesar et al. 2017). The mask is the same in both panels and is based on the PS1 Dec limit, masking the plane and the two prominent arms of Sagittarius at these distances (PMs are too noisy for RRL at these distances, precluding the possibility of using PMs to select Sgr stars as we have done for the K giants). Notice the strong detection of the local wake, as well as the broad* *overdensity in the North visible in the PS1 data. These features are nearly absent in Gaia data, indicating that the latter data cannot be used to probe the outer halo.*

Reviewer Reports on the Second Revision:

Referees' comments:

Referee #1 (Remarks to the Author):

I thank the authors for their additional work, which has strengthen the paper. There remain a few minor issues that would need further clarification before I can recommend this manuscript for publication.

* I do not agree with the conclusion that the tilted halo model is a worse fit to the data, given that both scenarios (i.e. misaligned halo in equilibrium vs. perturbed spherical halo) show clear discrepancies with the data (e.g LMC simulations overpredicts the density of halo stars at l>0 b>0, whereas the tilted halo underpredicts the density in that region). Without a more quantitative analysis this referee cannot tell which fits better. It is also not clear whether the current comparison with data is done in a similar way as with the LMC model (i.e. move the observer to the solar location).
It seems to me that kinematic infomation will be key to distinguish between equilibrium/disequilibrium models (see also my point below).

This paper would benefit from a discussion of the above results, and I encourage the authors to add the comparison with a tilted halo in the rebuttal as Extended Data.

* This paper claims that the detection of the local wake and Northern over-density in the outskirts of the stellar halo conforms with the predictions from N-body models where a massive LMC is falling in for the first time.
The authors may be aware of another independent prediction that has been recently validated by observations, namely that the LMC pull should dislodge the Milky Way disc from the barycentre of the Galaxy. Petersen & Penarrubia (2020, NatAs, 236) have recently detected the disc motion imprinted in the kinematics of distant K Giants and BHBs with available radial velocities and Gaia proper motions.
It thus appears that both the distribution & kinematics of the distant halo stars show clear disequilibrium features consistent with the predictions from LMC models accreted recently on to the Milky Way.
My suggestion is to highlight the agreement between the two independent analysis in the text, which in turn disfavours the tilted-halo/equilibrium scenario.

* The following sentence is not entirely correct: "A large-scale radial velocity asymmetry in the outer halo has recently been reported and has been interpreted as being due to the LMC^{32, 33}. The transverse velocity signal is expected to be of similar amplitude and may be detectable with either the recent or future data releases from Gaia. " Note that Petersen & Penarrubia (2020) use K Giants, BHBs and satellite galaxies with available Gaia DR2 proper motions to measure the disc motion with respect to the outer halo. These authors show that the availability of transverse motions is crucial to distinguish a displacement of the Milky Way disc from rotation in the halo. In contrast, Erkal et al. (2020) only used radial velocities in their analysis.

Referee #2 (Remarks to the Author):

I read again with interest the new version of "Detection of the All-Sky Response of the Galactic Halo to the Magellanic Clouds", by C. Conroy and collaborators.

Each new version of this article strengthens part of the argument, yet also raises new concerns. However, I feel that the authors are entitled to their interpretation of the data, and once published the community will do its job of testing the predictions of this analysis. It is likely to generate a lively debate.

Given the new analysis, I checked the distribution of Gaia DR2 RRLyrae, using a metallicity-distance calibration based on Gaia's fitted Fourier coefficients, and also using an alternative measure assuming $M_G=0.7$. Curiously, I found that the resulting [60-100]kpc samples show no sign of the overdensity reported in Figure 7 of the ED. Admittedly, this may be due to inhomogeneities in the Gaia DR2 RRLyrae catalog, but given the strength of the reported signal, I find this somewhat suspicious. (I only mention this in passing, and do not expect the authors to respond to this point, unless they wish to).

Regarding my question of the smoothed maps, these are obviously not masked and then smoothed. Given that the authors insist that this is the case, I presume they actually mean that they are masked, then smoothed, then masked again? If the final step is not a masking, the counts would be seen to "bleed" into the masked regions.

Author Rebuttals to Second Revision:

We thank the referees for additional helpful comments. Our replies to them are included below.

Sincerely
Charlie Conroy on behalf of the authors.

Referee #1 (Remarks to the Author):

* I do not agree with the conclusion that the tilted halo model is a worse fit to the data, given that both scenarios (i.e. misaligned halo in equilibrium vs. perturbed spherical halo) show clear discrepancies with the data (e.g LMC simulations overpredicts the density of halo stars at $l>0$ $b>0$, whereas the tilted halo underpredicts the density in that region). Without a more quantitative analysis this referee cannot tell which fits better. It is also not clear whether the current comparison with data is done in a similar way as with the LMC model (i.e. move the observer to the solar location). It seems to me that kinematic information will be key to distinguish between equilibrium/disequilibrium models (see also my point below).

This paper would benefit from a discussion of the above results, and I encourage the authors to add the comparison with a tilted halo in the rebuttal as Extended Data.

REPLY:

In response to the referee's comment, we have added a new section to the Methods ("A Tilted Halo?") and a new ED Fig 9. We note that the prolate tilted halo model is indeed treated exactly like the data (e.g., the observer is at the solar location, same masking, smoothing, etc.)

* This paper claims that the detection of the local wake and Northern over-density in the outskirts of the stellar halo conforms with the predictions from N-body models where a massive LMC is falling in for the first time. The authors may be aware of another independent prediction that has been recently validated by observations, namely that the LMC pull should dislodge the Milky Way disc from the barycentre of the Galaxy. Petersen & Penarrubia (2020, NatAs, 236) have recently detected the disc motion imprinted in the kinematics of distant K Giants and BHBs with available radial velocities and Gaia proper motions. It thus appears that both the distribution & kinematics of the distant halo stars show clear disequilibrium features consistent with the predictions from LMC models accreted recently on to the Milky Way. My suggestion is to highlight the agreement between the two independent analyses in the text, which in turn disfavours the tilted-halo/equilibrium scenario.

REPLY:

Yes, we are aware of this paper (it was included and cited in the last referee round), and agree that it disfavors the tilted halo solution. We have explicitly noted this in the new Methods section ("A Tilted Halo?"), and have added text in the final paragraph of the main paper highlighting the fact that these are two independent approaches coming to the same conclusion (see below).

* The following sentence is not entirely correct: "A large-scale radial velocity asymmetry in the outer halo has recently been reported and has been interpreted as being due to the LMC^{32, 33}. The transverse velocity signal is expected to be of similar amplitude and may be detectable with either the recent or future data releases from Gaia. " Note that Petersen & Penarrubia (2020) use K Giants, BHBs and satellite galaxies with available Gaia DR2 proper motions to measure the disc motion with respect to the outer halo. These authors show that the availability of transverse motions is crucial to distinguish a displacement of the Milky Way disc from rotation in the halo. In contrast, Erkal et al. (2020) only used radial velocities in their analysis.

REPLY:

We thank the referee for catching this error. We have updated the final paragraph to read:

Simulations that include the effect of the LMC on the Galactic halo\cite{Garavito-Camargo19, Cunningham20, Petersen20} predict large radial and tangential velocity differences across the sky due to the dynamical response of the halo to the LMC, with amplitudes exceeding $45\, {\rm km\,s}^{-1}$. The sample presented here will be ideal for measuring the velocity signal from spectroscopy and proper motions, as nearly all stars are brighter than $G=17.5$. Independent work has recently detected the predicted velocity signatures\cite{Petersen20b, Erkal20b}.

These detections provide strong corroboration that the density variations reported here are due to the dynamical response of the Galaxy to the presence of the Magellanic Clouds. The joint mapping and modeling of the density and kinematic signatures is the next step in uncovering the complex phase space structure imprinted by the Magellanic Clouds in the outer Galaxy.

Referee #2 (Remarks to the Author):

Given the new analysis, I checked the distribution of Gaia DR2 RRLyrae, using a metallicity-distance calibration based on Gaia's fitted Fourier coefficients, and also using an alternative measure assuming M_G=0.7. Curiously, I found that the resulting [60-100]kpc

samples show no sign of the overdensity reported in Figure 7 of the ED. Admittedly, this may be due to inhomogeneities in the Gaia DR2 RRLyrae catalog, but given the strength of the reported signal, I find this somewhat suspicious. (I only mention this in passing, and do not expect the authors to respond to this point, unless they wish to).

REPLY:

We note that the lack of a detection in Gaia RRL was reported in our previous reply to the referees (see Supp. Fig 2). Furthermore, we have stated this explicitly in the RR Lyrae Section in Methods:

Further progress could be made employing an all-sky map of RR Lyrae. {\it Gaia} data release 2 provided such a sample\cite{Rimoldini19, Clementini19}. However, these maps are incomplete at the depths necessary to detect RR Lyrae at >60 kpc ($G\approx20$), driven largely by the {\it Gaia} scanning pattern\cite{Mateu20}. For example, the local wake is not clearly visible in the {\it Gaia} RR Lyrae catalog despite the fact that it is so prominent in the Pan-STARRS data at >60 kpc. Future data releases from {\it Gaia} will hopefully reach the depths necessary to deliver complete samples of RR Lyrae in the outer halo.

In short, we agree completely with the referee that the features identified in the K giant sample are not visible in the Gaia RRL sample, though they are visible in the Pan-STARRS RRL sample.

Regarding my question of the smoothed maps, these are obviously not masked and then smoothed. Given that the authors insist that this is the case, I presume they actually mean that they are masked, then smoothed, then masked again? If the final step is not a masking, the counts would be seen to "bleed" into the masked regions.

REPLY:

The referee is correct - the healpy smoothing program automatically replaces masked regions (which are assigned special value) with 0.0s in order to perform smoothing. The masked regions are then automatically returned to the special values when the array is returned. So while we were not explicitly doing: mask, smooth, mask, the internal workings of the code was performing exactly this logic. We apologize for the confusion. We have added the following sentence to the Methods to further clarify this point: “Finally, the mask is re-applied to the maps after smoothing to avoid the appearance of the data ``bleeding'' into the masked regions.”