The comparison between the total flow measurements and these two calculations illustrates a qualitative agreement. This agreement between the data and the models depends on the particle species, transverse momentum range and centrality percentile and overall the AMPT model reproduces these measurements more accurately than TRENTo [19].” ==> we need to go beyond the “qualitative nature” of this comparison; I feel we need to quantify i.e. are these models working better/worse for the total than the NL modes? And by how much? What does this tell us? Are models capable of describing the total flow coefficients better than the NL terms? If yes, we could probably make an additional point?

Answer: You're absolutely right on the fact that the comparison should be quantified. So we tried to quantify the performance of these models in reproducing the measurements of total flow wrt non-linear flow modes by

1) taking the χ2/N_dof between each model and the measurement

2) fitting the relative ratios between the models and the data (both v_n and v_n,mk)

These exercises are included in the following twiki link:

https://twiki.cern.ch/twiki/bin/view/ALICE/Nonlinearflowvsmodels
https://twiki.cern.ch/twiki/bin/view/ALICE/PublishedTotalFlowMeasurements

Our interpretation of the compatibility of different models with the measurements as included in the new version of the paper:

"These two models have been utilised before to reproduce the p_T-differential v_n measurements for identified particles [40]. In order to compare the performance of these two models in v_n and $v_{n,mk}$ measurements, the relative ratios between each model and the measurements have been obtained. Tables 3 and 4 summarize these relative ratios for $v_{n,mk}$ and v_n, respectively. The ranges in the tables present the minimum and maximum value of a constant fit to the relative ratios obtained from most-central to mid-peripheral collisions. These values should be taken with caution as the non-linear flow modes have smaller magnitude and any discrepancy between the models and the data becomes magnified in the ratios. Comparison between Tab. 3 and 4 shows that the AMPT calculations reproduces $v_{4;22}$ with 20% higher discrepancy on average compared to v_4, while, TRENTo calculations performs better in $v_{4;22}$ compared to v_4 with 7%.

All in all, this study shows larger discrepancy between the model calculations and $v_{n,mk}$ measurements wrt. that of v_n, indicating a larger sensitivity to the initial conditions and transport properties in non-linear flow modes. As a result, it is useful to tune the input parameters of hydrodynamic models using the non-linear flow measurements and constrain the values of transport properties and the initial conditions of the system."

- **Along the previous lines but more importantly since this is related to the data points, we need to be a bit more systematic in the way we report the main features of the NL terms i.e. mass ordering and the particle-type grouping and try to go one step further from just the observation. ==> Is this mass ordering more pronounced for the NL terms than the total flow? Maybe not (at least calculating things by eye), but please quantify! IS the particle type grouping holding better for the NL terms than the total flow? Maybe not (at least calculating things by eye), but please quantify!**

Answer: Again you're absolutely right on this. The comparisons were only qualitative. So I quantified the difference in the character features of v_422 and v_4 (only the fourth harmonic as we don't have measurements of v_5 with run2 and the run1 measurements are lacking statistics for a proper comparison). The comparisons and the difference between mass ordering as well as particle type grouping and the fits in the two pT regions where these features are seen can be found in the following twiki link:

https://twiki.cern.ch/twiki/bin/view/ALICE/V4vsV422

These comparisons overall, show no apparent difference between the anisotropic flow and the non-linear flow modes in terms of mass ordering or particle type grouping which is also reflected in the plots in: https://twiki.cern.ch/twiki/bin/view/ALICE/V4vsV422

In particular comparisons for mass ordering both via taking a difference and ratios are shown in :

difference: https://twiki.cern.ch/twiki/pub/ALICE/V4vsV422/
MassorderingDifferencev422v4_firsthalf_allcentralities.pdf
ratio: https://twiki.cern.ch/twiki/pub/ALICE/V4vsV422/
MassorderingRatiov422v4_firsthalf_allcentralities.pdf

The differences are compatible with 0 within 1σ. The ratios are compatible with 1 within 1 or 2σ.

comparisons for particle type grouping:
difference: https://twiki.cern.ch/twiki/pub/ALICE/V4vsV422/
MassorderingDifferencev422v4_secondhalf_allcentralities.pdf
ratio: https://twiki.cern.ch/twiki/pub/ALICE/V4vsV422/
MassorderingRatiov422v4_secondhalf_allcentralities.pdf

The differences are compatible with 0 within 1 or 2σ and the ratios are compatible with 1 within 1 or 2σ.

We also performed the same exercise for the model calculations by comparing their predictions in v422 and v4.
Here is the link to these comparisons, differences and ratios:

https://twiki.cern.ch/twiki/bin/view/ALICE/V4vsV422

Where figure 11 and 12 show the comparisons, 13 and 14 show the differences and 15 and 16 show the ratios of the mass orderings in v422 and v4.

As a result, we added a subsection to the paper mentioning that the mass ordering and particle type grouping in the non-linear modes is compatible with that seen in vn measurements.

"The features seen in the measurement of non-linear flow modes can be further studied by comparing to that of total flow coefficients. Such comparisons have been performed for v4,22 (pT) (this study) and v4(pT) measurements [40] by taking the relative difference of pions wrt protons at a given pT in both modes. This comparison shows that the observed mass ordering in low pT region ($0 < pT < 2.5$ GeV/c) is of the same magnitude in v4,22 and v4. In the intermediate pT region ($pT > 2.5$ GeV/c) the observed particle type grouping also shows the same magnitude in both flow modes."

Please let us know if these lines are not sufficient and more in depth discussion is needed and/or they should be revised.