Response to Reviewers

Reviewer 2

We thank Reviewer #2 for their comments and suggestions on our manuscript "Inverse modelling of New Zealand's carbon dioxide balance estimates a larger than expected carbon sink". In the text below, we have included all the original reviewer comments and suggestions in black, followed by our response and relevant manuscript changes in red. Page and line numbers refer to marked-up manuscript version.

This reviewer feels this paper is in scope for ACP, presenting atmospheric top-down carbon dioxide estimates across New Zealand, using two in-situ sites (North and South island) across a 10-year period. The key result for the paper is a larger carbon sink than previously reported compared with both bottom-up estimates and from the study this paper builds on. The potential sources for this difference are presented and discussed. This also presents an updated biospheric flux model tailored to the region.

This paper is thorough and well written, includes interesting and well-justified conclusions, has appropriate references to current literature throughout and some novel work which could prove useful for the community. The scientific methods are well discussed, with details given on the sets of inputs to the inversion model. Some details for the inversion model itself are included in the Supplementary Information but as this is the focus for the paper these methods should be included and discussed in the main text (as also noted and expanded upon by Reviewer 1).

We thank Reviewer #2 for their careful revision of our paper and feedback. As suggested by both Reviewer #1 and #2 we have moved the inversion description from the Supplementary Information (S1) to the main paper (now Section 2.1)

Specific comments below:

L153: "Our inversion used 4-day integrated air concentration (i.e., footprints, units g s m-3, Fig. 3), averaged throughout the Planetary Boundary Layer (PBL)". What is meant by averaged throughout the Planetary Boundary Layer (PBL) here? Was this taken at various heights above the surface based on the PBL values in NAME meaning this was variable for each footprint?

That is correct, each footprint (i.e., the vertical extent of the footprints) will vary with each grid box and time step depending on the height of the modelled PBL. The boundary layer height field is available to NAME III on model levels (70 levels ranging from the surface to 40km above the surface). Any averaging is done from the surface up to the diagnosed boundary layer height over the requisite number of levels. A detailed description is available in Section 1.4.1.1 and Table 1.18 in https://code.metoffice.gov.uk/doc/name/vn7.2/docs/namedoc_A01.pdf (accessing the document requires registration with MetOffice).

L159: "By the end of the 4 days, most of the particles had left the model domain." Has this been tested for the footprints involved in this study as well as Steinkamp et al., 2017? A 4-day period seems short but is this due to small domain and large wind speeds?

This has been tested both in Steinkamp et al., (2017) and for this study, although the results are not shown in the paper. NAME III provides a specific diagnostic that tracks how many particles

leave the modelled domain. Figure 1 shows an example of the analysis for year 2014 for both Baring Head and Lauder.

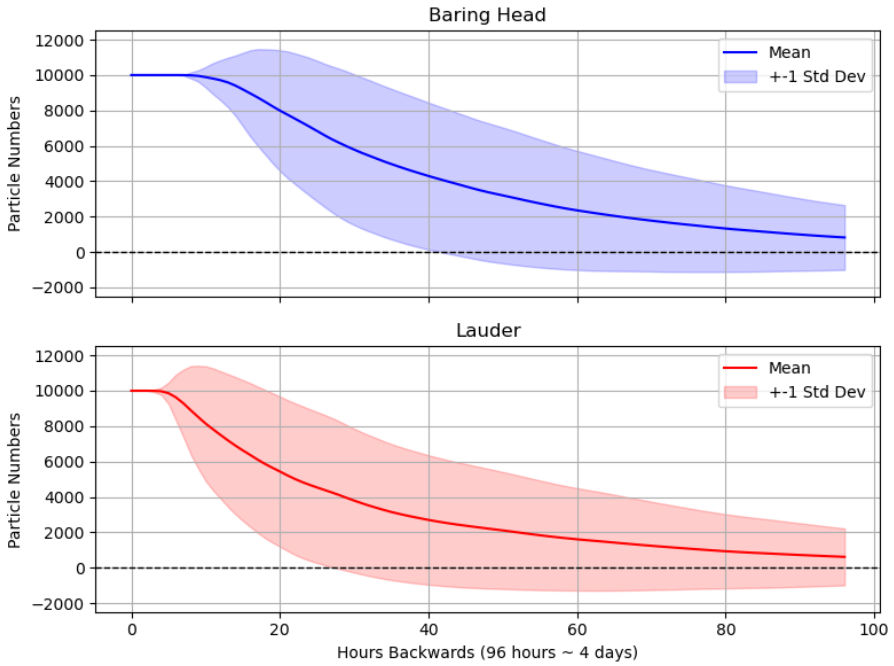

Figure 1. Mean and standard deviation of particle counts over time for Baring Head (top) and Lauder (bottom) from backward trajectory simulations over 96 hours (~4 days). The solid lines show the mean number of particles remaining, while the shaded regions represent ±1 standard deviation. Particles are released at hour 0, with numbers decreasing as they exit the domain. The dashed line at y = 0 marks complete particle loss.

For each measurement time, we release 10,000 particles (y-axis), and the majority of the particles leave the domain after 4-days (i.e., 96 hours, ~0 particles). Using 2014 as an example, for Baring Head on 80% of the days more than 90% of the particles left the domain, while for Lauder on 85% of the days more than 90% of the particles left the domain. The particles that remain in the domain are during days with very low wind speed. Other years follow the same pattern.

Section 2.2.1 Atmopsheric model transport validation: Why are wind speed and wind direction the most important meteorlogical parameters to investigate here? Were there other parameters which may be significant as well?

We appreciate the reviewer's comment and the opportunity to clarify our choice of meteorological parameters for validation. Wind speed and wind direction are the primary drivers of transport in NAME III, directly influencing the dispersion and advection of atmospheric tracers. Since NAME III is a Lagrangian transport model, accurate representation of these parameters is critical for ensuring that air parcels are correctly advected from source regions to measurement sites. In addition to wind fields, we also evaluated planetary boundary layer (PBL) heights at Lauder (Figure S5), recognizing that vertical mixing is another key factor affecting particle transport. However, the validation was limited by the availability of verification data. We believe that wind speed, wind direction, and PBL height represent the most essential diagnostics for evaluating transport accuracy in this context. Furthermore, we provide a comparison of the topographical representation in the model (Figure S4), demonstrating the improvements achieved with higher-resolution meteorological input data (i.e., NZCSM).

Section 2.4 Prior and posterior uncertainties: Would be good to see justification for the uncertainty terms chosen including:

- "we fixed (i.e. summed) 50% of the uncertainty term"

- "The uncertainity for pine fluxes from CenW were assumed to be 30%, 30%, 60% of the NEE, GPP and ER flux magnitudes"

- what was the justification for these choices for uncertainties? Do we know they are representative (or that the extremes included within the sensitvity tests are justified)?

The fixed (i.e., summed) 50% uncertainty term was chosen to ensure realistic posterior uncertainties, as lower values appeared small given known transport and model-data mismatch errors. This adjustment accounts for systematic uncertainties and underrepresented variability, preventing overconfidence in the estimates. Sensitivity tests confirm that our approach captures a reasonable range of uncertainty while avoiding extreme underestimation (Figure 15, Page 23).

In terms of the pine fluxes from CenW, the model had been parameterised against growth data from 1309 observations from 101 stands located throughout New Zealand (Kirschbaum & Watt, 2011). That data set covered stands growing under all the environmental conditions over which *Pinus radiata* is grown in New Zealand. Comparison between observed and modelled volume growth was described with a Nash-Sutcliffe model efficiency of 0.853, hence providing high confidence in the model's performance in describing the overall long-term growth potential (Kirschbaum & Watt, 2011).

Translating that into confidence of short-term NEE fluxes adds some additional temporal and spatial uncertainty, although both were constrained by the broad coverage in the observations. Hence, an NEE uncertainty of 30% seemed conservative given the model efficiency of 0.853 in the underlying model-data comparison. As the major flux driving NEE is GPP, and, given that part of the uncertainty in NEE results from uncertainty in ER, we assigned the same uncertainty to GPP as had been determined for NEE.

For ER there are two additional main sources of uncertainty:

1) The proportion of GPP captured in ultimate growth as reflected in NEE can vary through physiological factors, such as the allocation to different biomass components, such as foliage and branches, that senesce eventually and are lost through respiration, versus allocation to stems that are retained in the measured eventually observed wood growth;
2) The respiratory contribution from the decomposition of organic matter or other plant residues. In this context for second and later-rotation pine stands, the decomposition of dead coarse roots from preceding rotations are particularly important through their quantitatively large contributions. We estimated that approximately ¼ of carbon fixed during the growth of a pine stand is allocated to the root system. Over the next rotation, that carbon will eventually be lost by respiration after the root system is killed during harvesting. Uncertainties relate to the timing of that loss, its magnitude and whether stands are first rotation or subsequent rotation stands.

Given these two factors, we estimated the uncertainty of ER fluxes to be twice as large as the uncertainty of GPP fluxes

References: Kirschbaum, M. U. F., & Watt, M. S. (2011). Use of a process-based model to describe spatial variation in Pinus radiata productivity in New Zealand. Forest Ecology and Management, 262(6), 1008-1019.

Table 2: Exclusion of Test 9 and 10 from the table itself (and details included in the caption) is a little confusing when comparing to Figure 14. Would be good to see the details of these tests included within the table.

Details of both Test 9 and 10 are now included in Table 2 (Page 24).

Supplementary Info

SL54: "NAME III vn6.5 dispersion model for the years 2014-2020 while for the period 2011-2013, we retained the original NAME III vn6.1 simulations" - was any comparison made between the outputs between the two NAME versions? Did this make any difference to the sensitivity profiles?

A comparison between the two NAME III versions was conducted for ~6 months in 2014 (January to July) when compatible input data were available for both models. The results showed minimal differences, with no significant impact on the sensitivity profiles. The core science and transport schemes relevant to the inversion remained unchanged between NAME III vn6.1 and vn6.5. The updates primarily involved technical modifications or changes to schemes that were not utilized in this study, ensuring consistency in our inversion framework.

SL68: "00Z" - does this mean UTC at midnight?

Yes, "00Z" refers to midnight 00:00 UTC and equivalent to 12 UTC in New Zealand, following a convention in Numerical Weather Prediction (NWP) to use UTC or "Z" as a consistent time reference for global coordination.

Figure S6: Consider updating the red-green in this plot to something colour-blind friendly (this could include changing one of the maker styles to be distinctive).

We thank the reviewer for this suggestion, we have updated the colours of the figure (and figure caption), designed to be distinguishable for people with colour vision deficiencies (now Figure S5, Page 7).

Table S3: For the "Area (m2)" column is the area itself known to this level of precision?

We appreciate the reviewer's comment. The high precision in the "Area (m^2)" column was a result of the table being generated automatically from the inversion output files, which contained area values with the specified level of precision. We recognize that such precision is unnecessary for reporting purposes and we have now removed the decimal values for clarity (Page 14).

Figure S17: "Diurnal cycle test results for the land inversion regions." Caption is not too descriptive and requires other details to understand the context for this. Please add a few more details here to this caption more standalone.

We have added additional information about the figure. Specifically: *Diurnal cycle test results (Section 4.3, main paper) for the land inversion regions. The figure shows the timeseries of prior (black) and posterior (green) fluxes for each land inversion region (i.e., region 1-15), along with corresponding annual values. The numbers above the annual plots represent the prior (top) and*

posterior (bottom) annual fluxes in units of Tg CO_2 yr^{-1}. Identical prior and posterior fluxes indicate no systematic bias in the inversion system, while any differences suggest a potential over- or underestimation of CO_2 sources or sinks. (now Figure S16, Page 31)

Typos and grammer:

- L42 scalMOLes --> scales Fixed (Line 42)

- L123-124 (Terao et al., 2011; Yamagishi et al., 2012; Müller et al., 2021)) - Bracket within a bracket here, this should be typeset to remove this. Fixed (Line 157)

- L229: Fore --> For Fixed (Line 265)

- L256: Front quote around 'other forests' is a back rather than a front quote Fixed (Line 292)

- L726: "the inverse model detect" --> grammer issue here Fixed (Line 767)

Supplementary Info

- SL68: creata --> create Fixed (Line 72)

- Figure S16: Tabel 3. --> Table Fixed (now Figure S15, Page 31)

- Figure S19: Timseries --> Timeseries Fixed (now Figure S18, Page 33)