Authors' responses to Referee Comments*, Critical Load Exceedances for North America and Europe using an Ensemble of Models and an Investigation of Causes for Environmental Impact Estimate Variability: An AQMEII4 Study,* **Referee's comments are in** *italics*, **responses in** regular font.

December 19, 2024.

We thank the Referee's for their feedback – we believe that the Referee's comments have been addressed and answered in depth. Our detailed responses follow.

Anonymous Referee # 1

General comment

- *This manuscript, "Critical Load Exceedances for North America and Europe using an Ensemble of Models and an Investigation of Causes for Environmental Impact Estimate Variability: An AQMEII4 Study" presented by Maker et al., summarized their excellent work in the model intercomparison study of AQMEII4. The authors reported a comprehensive investigation regarding the model variability and suggested potential research directions for future studies. I believe that this kind of study is crucial to progressing our knowledge of the model themselves and their applications. Most of my comments are minor specific and technical comments for better reading and presentation quality; however, I would like to make one major request regarding the current manuscript.*

We thank the reviewer for these words. This has been a big effort; a four-year project by 27 collaborators, in part delayed by the Pandemic. Very nice to hear it's appreciated!

Major comment

- *As described in Section 2.0, model simulations over Europe were carried out in 2009 and 2010, which is why there is a large difference in the meteorological field. From this manuscript, I can follow the discussion in the U.S., which targeted a significant SO2 emission reduction between 2010 and 2016. However, from the result over Europe such as shown in Figs. 3 and 5 and Table 4, the difference between the years 2009 and 2010 was small. This discussion could be presented in other companion papers in AQMEII4 project, but one possible conclusion from this manuscript is that there is little impact on the estimates of critical loads despite the significant impact from the meteorological field? Actually, I did not fully figure out what was the most variated parameters (temperature, wind field, precipitation, etc.) between these years. For more information derived from this study, please consider including this point.*

The Referee raises a good point here – while we mentioned that the reasoning for our choosing the years 2009 and 2010 was the differences in the weather between the years (as well as the availability of observation data for model evaluation in both years), we did not elaborate on the weather component. As the referee points out, the differences in the weather between the two years clearly had a relatively small impact on the deposition fluxes and hence on the critical load exceedances.

To address the Referee’s comment, additional text has been added to the description of the model simulations; we have replaced the sentence “European years were chosen due to a large difference in meteorology between 2009 and 2010, hence allowing the effects of potential meteorological on deposition to be estimated..” with “The European years were chosen due to a large difference in meteorology between the years 2009 and 2010, the latter being a year with unusually high summer temperatures eastern Europe and the western side of the Russian Federation (Barriopedro et al., 2011) leading to increased European forest fire activity and emissions during that year (Schmuck et al., 2011). The July 2009 and July 2010 temperature anomalies relative to the base year period 1961 to 1990 are shown in Supplemental Information Figure S1). The precipitation anomalies in July of each year are less significantly different than the temperature anomalies; similarly, the differences between the annual average temperature and precipitation anomalies between the two years is less significant than the July values. In the analysis which follows, the differences in simulated deposition and critical load exceedances for European region between the two years is shown to be relatively minor, implying that forest fire emissions contributed a relatively small proportion of sulphur and nitrogen deposition in 2010, and that the summer temperature anomalies in 2010 did not result in significant perturbations to total sulphur and nitrogen deposition.” Note that we have also mentioned elsewhere in the originally submitted manuscript that the differences between the anthropogenic emissions between the two years is not significant.

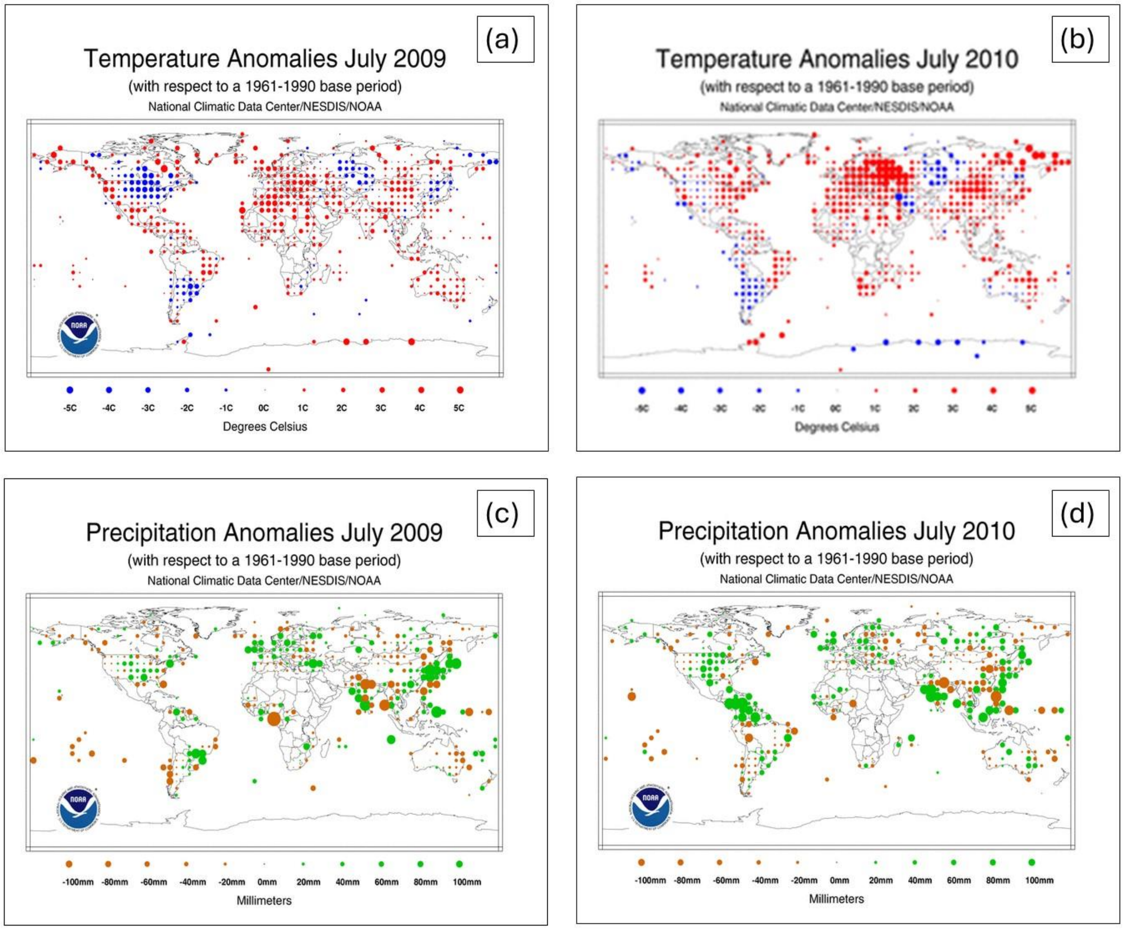
Additional references:

Barriopedro, D., Fischer, E.M., Luterbacher, J., Trigo, R.M., and Garcia-Herrera, R., The hot summer of 2010: redrawing the temperature record map of Europe, *Science*, 332, 220-224, 2011.

Schmuck, G., San-Miguel-Ayanz, J., Camia, A., Durrant, T., Santos de Oliviero, S., Boca, R., Whitmore, C.J., Giovando, C., Liberta’, G., Corti, P., Schulte, E., Forest Fires in Europe 2010. EUR 24910 EN. Luxembourg (Luxembourg): Publications Office of the European Union; 2011. JRC66167. https://publications.jrc.ec.europa.eu/repository/handle/JRC66167

New Supplemental Figure S1:

Figure S1. Temperature (a,b) and precipitation (c,d) anomalies relative to the 30-year period 1961-1990, for the years 2009 (a,c) and 2010 (b,d). Note the large positive anomaly (red colours) in temperature for July of 2010 over Europe (b). Data and images from NOAA National Climatic Data Center, https://www.ncei.noaa.gov/access/monitoring/ghcn-gridded-products/maps/, last accessed November 19, 2024.



Specific comments

- *L37 and L40: In the abstract, "New" is repeated, and this wording will be ambiguous. It will be better to use another specific term.*

"New, targeted" has been changed to "Targeted", and New datasets for North American critical loads for acidity for forest soil water and aquatic ecosystems were combined…" has been changed to "Datasets for North American critical loads for acidity for forest soil water and aquatic ecosystems were created for this analysis. These were combined…".

- *L157-160: As written in L473-475, it is better to mention that this time is the year 2021 status explicitly.*

The sentence has been changed to end "at the time the simulations and critical load data collection took place (2021)."

- *L177: For modelers, "process analysis" will be associated with the model-embedded process analysis tool (https://www.cmascenter.org/cmaq/science_documentation/pdf/ch16.pdf). Could you change the wording here?*

While noting that the process analysis tool mentioned may also be new to some readers of the manuscript (including the first author!) – to avoid potential confusion with that tool, the sentence has been changed to "The work conducted here uses analysis of new model diagnostic outputs added for AQMEII-4 to attempt to determine the key causes of these model deposition estimate differences."

- *L198: In the latter part of the discussion, we can follow the wording "reduced ensemble", but this term is suddenly used here without any introduction. Please rewrite or define this wording here.*

Thanks – in this particular line, the "reduced ensemble" refers to that of Vivanco et al. (2018); we've modified the text to better introduce the idea behind the term, which helps put our later use of the same term in context. The two sentences now read "The models with the best performance relative to observations were used to provide ensemble critical loads – a "reduced ensemble" in that not all models submitting output for the study were used in generating ensemble critical loads. However, even within this reduced ensemble, local variations of over a factor of four in both sulphur and nitrogen deposition could be seen between the ensemble members, and the predicted percent area in exceedance for sensitive ecosystems varied by more than a factor of two for the best performing models (Vivanco et al., 2018)."

- *L416: What is "pant" here? Is this a typo of "plant"?*

Oops! Yes, that should be "plant"! Corrected.

- *L413-415: "A second model..." is described in L415, so "A first model" can be explicitly stated in L413?*

Sure. Done. Line 413 sentence now starts with "A first model was developed".

- *L423: For wide readers, "PRISM" should be shortly introduced.*

Good point. We should have also mentioned that PRISM refers to the interpolation model used. The segment of the sentence now reads: "and Parameter-elevation Relationships on Independent Slopes Model (PRISM) interpolation data for temperature and precipitation (Daly et al., 2008),"

- *L476-477: Because the following sentences started from the U.S. and then stated Europe, it is better to change the position here (i.e., the years 2010 and 2016 for North America, and 2009 and 2010 for the European region).*

Ok. First sentence of that paragraph now has them reversed as suggested, “Model simulations were carried out for the years 2010 and 2016 for North America, and 2009 and 2010 for the European region.”

- *L549: This section introduced participating models. We can follow the result section (L708) from these descriptions, but I think it would be helpful for readers to summarize in a table which model was performed for which or both domains (Northern America and Europe).*

Rather than an additional Table, we’ve added this information in a sentence in the text at the start of section 2.2, as follows, “The models CMAQ-M3Dry, CMAQ-STAGE, WRF-Chem (IASS), GEM-MACH (Base), GEM-MACH (Zhang), GEM-MACH (Ops), WRF-Chem (UPM), and WRF-Chem (UCAR) provided simulations for AQMEII-4, interpolated to the common the North American domain. The models WRF-Chem (IASS), LOTOS-EUROS (TNO), WRF-Chem (UPM) and CMAQ (Hertfordshire) provided simulations for AQMEII-4, interpolated to the common European domain. Some of the modelling frameworks were repeated, but process implementation details were varied in order for the relative impact of these differences to be examined. We describe each of these models according to the starting framework (CMAQ, GEM-MACH, WRF-Chem, LOTOS-EUROS), below.”

- *L961 (Fig. 14), L1397 (Fig. 22), L1422 (Fig. 23), L1490 (Fig. 24), L1667 (Fig. 27), and L1740 (Fig. 30): The gray grid indicated the negative value from the color bars; however, I think this is just out of the calculated domain.*

Correct: the value of ‘-9’ was used as a field mask to indicate “outside of the common domain” in these figures. A note to this effect has been added to each Figure caption in the revised manuscript, “Note that regions outside the common AQMEII-4 domain have been assigned an “outside domain” mask value of -9.” We updated Figure captions to ensure that all had the same font size in the process.

- *L1575 (Fig. 26): Because the color scale is based on blue-red bars, it is tough to distinguish blue and red lines, which indicate the predominant land use category. Please revise this figure.*

Done – we’ve used a green line for the agricultural plus grasslands land use type, and a purple line for the forested land use type, which make them both more visible against the background colours.

Technical corrections

- *L98: Please correct “A a”.*

  Done.

- *L119: “simple mass balance (SMB) model” will be better.*

  Done.

- *L224: No need “)”.*

Done.

- *L236, L239, and L240: Please use subscript for "PM2.5" and "PM10".*

  Ok – I see that this is the standard being used *in Atmospheric Chemistry and Physics*, from a quick look at other papers already on-line. This seems to depend on the journal (e.g. Atmospheric Environment does not use a subscript form, Nature does, Science does).

- *L335 (Table 1): Maybe there is no need for parenthesis in "source" description*.

  Yes, ok; brackets modified for the Source column in Table 1 to just be around the dates.

- *L342, L339, L371, L393, and L408: For the consistent expression of the subsection name in L429 and L447, it could be used ":" like, "1.2.1 Canada: Aquatic Ecosystem Data".*

  Done.

- *L363 and L364: The charge for each ion should be presented*.

  We're glad you brought this up, since it reminded us of a convention that's commonly used for critical load exceedance calculation equations with which readers unfamiliar with the field may be unaware: all of the charge balance equations are in units of charge equivalents (i.e. moles of ion x number of charges in the ion). Ionic charge values are therefore not included in the equations – but we've included a bracketed comment in the revised text to this effect prior to equation 7 to make this clear to the readers, with the revised sentence now reading: "Where the lake acid neutralizing capacity [ANC]_limit is defined as the excess equivalents of cations – anions in lakewater (note that all quantities in these equations are in units of charge equivalents; number of moles multiplied by the charge of the ion, so by convention, charges are not included in the variable names in the exceedance formulae): "

Anonymous Referee # 2, October 24, 2024

Through inter-comparison among models and comparison with monitoring results, the performance of CTM models for modeling S and N deposition in US and Europe was systematically evaluated. The critical load exceedance was further estimated based on the ensemble deposition simulation. It is of great importance that the future improvement of the models was suggested.

We thank the reviewer for that rating of the work.

However, the paper is not a research report and is limited in length. Although critical loads were needed to calculate the exceedance by S and N deposition, the calculation method of critical load need not be detailed described in the main text, because only existing critical load databases were used in this study. Instead, relevant literature (or supplement) can be referred. There were too many figures and tables on the comparisons of modeling results in the paper. I suggest to make further summary and move most of them to the supplement.

The length of the paper was a concern of ours as well – while it was a joint effort by 18 different research organizations and 27 co-authors, the result was a very large paper. It's important for the critical load calculation methods to be included as part of the work in some form, since the implementation details of these can affect the resulting critical load and critical load exceedance estimates. However, we agree that the main focus of the work is to describe the exceedances themselves, and then delve in detail into the causes of model variability in those exceedance estimates. Consequently, we have made a brief summary (below) of the critical load methodology in the main body of the paper, and have moved most of the original text on this topic to Supplement, as requested by the Referee.

Revised/reduced CL description in main body of the paper:

"A brief summary of the six CL datasets used in this work is provided here – full descriptions of the methodology used to create the CL data are provided in the Supplement, section 1.

North American CL estimates for *acidity in forest groundwater* were generated using the Simple Mass Balance model (Sverdrup & Warfvinge, 1990; Sverdrup & De Vries, 1994), employing data from several studies within the U.S. and Canada (McNulty et al., 2007, 2013; Duarte et al., 2011, 2013; Phelan et al., 2014, 2016; Sullivan, 2011; Sullivan et al., 2012; Cathcart et al., 2024) Table S1 (Supplement) methodological information for these studies, such as the horizontal spatial resolution, dataset extent, plant-species-specific critical base cation to aluminum soil water ratio values, the approaches used to estimate soil base cation weather rates, losses of (non-sodium) base cations from the ecosystem through uptake via harvesting or grazing, and whether nitrogen uptake via harvesting/grazing was included in the calculation of nitrogen minimum critical loads.

The North American *Aquatic Ecosystem acidity* critical load dataset constructed here combined individual datasets from the Canada and the USA, as follows.

Environment and Climate Change Canada data corresponding to the subset of 2,997 lake surveys which reside within the common AQMEII4 North American grid were used in conjunction with the Steady-State Water Chemistry (SSWC) critical load model (Sverdrup *et al.,* 1990) as described in Aherne and Jeffries (2015). SSWC is in widespread use for aquatic ecosystem CL

(Posch *et al.,* 2001; Cathcart *et al.,* 2016; Henriksen *et al.,* 2002; Jeffries *et al.,* 2010; Scott *et al.,* 2010; Whitfield *et al.,* 2006; Williston *et al.,* 2016; Dupont *et al.,* 2005; Miller, 2011). CL calculations for Canada followed a hierarchy based on the available information for individual lakes (for example catchment runoff rates were determined by isotope mass balance estimates in preference to a GIS map based approach using regional datasets, and when dissolved organic carbon estimates were available, an organic acid adjusted limiting value of the acid neutralizing capacity was used to include the influence of organic acids in the lake in preference to a fixed value of 40 μeq L^{-1}. Only sulphur deposition was used to determine exceedance, since the SSWC model does not consider non-acidifying nitrogen.

Aquatic ecosystem critical loads for the USA were taken from the National Critical Loads Database Version 3.2.1 (NCLDv3.2.1, Lynch *et al.,* 2022), which contains both the critical load data used here and supporting information. A total of 21,667 critical loads were used for 14,334 unique lakes and streams across the USA (a combination of different methods for determining the critical loads were included in the USA values, sometimes resulting in more than one CL estimate for the same water body). Most USA aquatic critical loads (78%) were determined using the SSWC model (Lynch *et al.,* 2022; Scheffe *et al.,* 2014; Dupont *et al.,* 2005, Miller 2011, VDEC (2003, 2004, 2012)), and site-specific catchment runoff rates (US EPA, 2023). The remaining 22% of USA aquatic critical loads were determined by a dynamic modelling approach (Sullivan *et al.,* 2005; Fakhraei *et al.,* 2014; Lawrence *et al.,* 2015) and a combination of dynamic modeling with a regionalization approach (McDonnell *et al.,* 2012, 2014; Sullivan *et al.,* 2012; and McDonnell *et al.,* 2021). Organic acid-adjusted *limiting acid neutralizing capacity* values were not used in generating these USA aquatic CL with respect to acidity datasets, and an average critical load value was used for these waterbodies for which overlapping CL estimates were available. A more detailed description of the USA aquatic critical loads used here can be found in Lynch *et al.,* (2022).

North American critical loads for *eutrophication* were estimated using CLE for two ecosystem types, sensitive epiphytic lichen, and herbaceous species richness.

CL for sensitive epiphytic lichen species richness made use of 9,000 community surveys across the USA from 1990-2012 (Geiser *et al.* 2019), where a 90% quantile regression was used to model relationships between deposition levels and observed species richness in order to estimate critical loads, and a -20% decline in species richness was used to determine the critical load. These methods resulted in a single critical load of 3.1 kg-N ha^{-1} yr^{-1} for sensitive epiphytic lichen, which was applied to all broadleaf, conifer, or mixed forest landcover types.

CL for USA herbaceous species richness made use of data developed using over 14,000 vegetation survey plots across nitrogen deposition gradients (Simkin *et al.,* 2016). An observation-based approach using median quantile regressions for herbaceous species richness response to deposition was employed, to generate critical loads with respect to nitrogen deposition linked to various atmospheric and soil conditions. Separate CL models were developed for open and closed canopies. The resulting CL of N for open canopy systems ranged from 6.2 to 12.3 kg-N $ha^{-1}yr^{-1}$ and the CLs of N for closed canopy systems ranged from 6.1 to 23.7 kg-N $ha^{-1}yr^{-1}$.

Two EU CL datasets were employed for the AQMEII4 EU domain, for *acidification* and *eutrophication* of *terrestrial ecosystems*, respectively. The critical load database and the exceedance calculations for Europe were provided by the Coordination Centre for Effects (CCE) under the United Nations Economic Commission for Europe Convention on Long-range Transboundary Air Pollution (UNECE LRTAP Convention), hosted by the Umweltbundesamt (UBA) in Germany, which develops and maintains the European critical loads database (Geupel *et al.*, 2022). The most recent database available was used here, and while country-dependent, all CL estimates made use of the Simple Mass Balance model (Sverdrup & De Vries, 1994; CLRTAP, 2023, Geupel *et al.*, 2022), with gap-filling using the CCE background database (Reinds *et al.*, 2021). Critical loads for EU eutrophication ($CL_{nut}N$) were also based on the SMB method applied to nitrogen deposition, and used two different methodologies to determine the accepted nitrogen leaching. Dependent on the country, empirical values were sometimes used as upper and lower boundaries for the SMB modelling results in order to avoid rather extreme results in ecosystems where the SMB model predicts very high or very low eutrophication CL values (Bobbink *et al.*, 2022). The resulting EU CLE were summarized as the share of the receptor area with critical load exceedance (bar charts) and the magnitude of the exceedance within each analysis grid cell (maps). The exceedance in a grid cell is defined as the so-called 'average accumulated exceedance' (AAE), which is calculated as the area-weighted average of the exceedances of the critical loads of all ecosystems in this grid cell."

In addition, the focus of this paper is on the uncertainty of critical load exceedance (CLE). Recognizing the uncertainty of deposition modeling, especially the underestimation of wet deposition, the degree of underestimation of CLE should be shown in the paper.

This is a good point. What we have done to respond to this is add another set of CLE calculations and images to the paper, for the year 2016 for North America and 2010 for Europe. For this final set of CLE estimates, the components of sulphur and nitrogen deposition for which observations are available have been bias-corrected across the model domain. That is, for sulphur and nitrogen chemical species for which air concentration observations are available, the ratio of mean observed concentration to mean model concentration across all observation stations generated, and this modelled dry deposition flux for that component was multiplied by this ratio (noting that the concentrations will be proportional to the deposition fluxes on average). Similarly the wet deposition fluxes have been bias corrected with the ratio of the mean observed to mean model wet deposition flux (a more direct correction). While not all of the model S and N species can be bias-corrected in this fashion (for example, there are no HNO_3 observations across North America with which to do a bias correction on HNO_3), we could nevertheless carry this out with the species for which observations *were* available. These bias-corrected model fields and the remaining S and N deposition fields (the latter were not bias corrected, due to lack of observation data) were then summed as before to generate total bias-corrected S and N deposition maps. These totals were in turn used to generate a third set of CLE, for the year 2016 for NA, and 2010 for EU, to quantitatively demonstrate the potential impact of the known model biases on CLE estimates. The results of the bias correction on North American and EU CLE has been added as an additional set of Figures in the Supplement (Revised Supplement, Section 4: Figures S9 to S14), and Figures 3, 5, 7, 9, 11 and 13 of the original manuscript (new manuscript Figures 2, 4, 6, 8, 10, and 12) have been updated to include an additional panel with the bias-corrected average CLE maps, and another with the bias-corrected values has been added to the corresponding bar charts. These allow the reader to quickly see an estimate of the impact of the model biases (when available) on the CLE values. A new section was added to the text describing this approach and contrasting it with more complex methodologies available in the literature. The excerpts from the revised paper dealing with this approach and the resulting conclusions follow:

"*2.3 Bias Corrected Critical Load Exceedance Estimates*

As will be discussed in Section 3.2, model results were evaluated using the available data for North America and Europe (see Supplemental, Section 7 for species contributing significantly to total S and N deposition). Critical load exceedances were calculated making use of the total sulphur and total nitrogen deposition for each model in the ensemble, for 2009 and 2010 for Europe, and 2010, 2016 for North America. In order to make a rough estimate of the impacts of model biases on the resulting exceedance estimates, a third set of exceedances were calculated for each model and each domain, for the year 2010 for Europe and 2016 for North America. For this last group, the ratio of the observed to model mean values at the observation station locations for individual species were used as scaling factors on the model annual deposition flux estimates prior to summation to total sulphur and total nitrogen deposition. Specifically, for North America, the ratio of the observed to measured mean concentrations of SO_2, NO_2, PM2.5 sulphate, PM2.5 ammonium, and AMoN network NH_3 were used to scale the corresponding dry flux variables, and the corresponding ratios for wet deposition of sulphate, nitrate and ammonium ions were used to scale the wet deposition fluxes. Less observation data were available for Europe than North America: the ratio of observed to modelled SO_2 and NO_2 gas concentration mean values were used to scale the corresponding dry fluxes, and ratios of observed to modelled wet deposition fluxes for sulphate, nitrate and ammonium were used to scale the modelled wet deposition fluxes.

We note that this approach makes simplifying assumptions. The corrections are inherently dependent on the assumption that the monitoring data is sufficiently representative of the model domain for the correction to be meaningful across the domain. While dry deposition fluxes will be proportional to the concentrations in the lowest model layer, allowing an overall mean bias correction, we are also making the assumption that the bias ratios for PM2.5 particulate matter will apply for larger particle sizes as well (note that size-resolved particulate fluxes were not reported under the AQMEII-4 protocol). This form of bias correction is also the simplest possible means of model-measurement fusion; more complex methods appear in the literature. These methodologies for example may make use of a combination of observed wet and adjusted model dry deposition (Schwede and Lear, 2014), inverse distance weighting from observation stations (Rubin et al., 2023) and adjusting modelled wet deposition fluxes by the ratio of observed to simulated precipitation and by kriged observed wet deposition to model predicted ratios (Zhang et al., 2019). An overview of model-measurement fusion approaches including advanced forms of data assimilation may be found in Fu et al., (2022). The methodology used here provides a first order estimate of the impact of model biases with respect to observations on critical load exceedances."

Additional text was added in the main body of the text with regards to the figures that were modified to include this information, in addition to the Figure captions being updated for additional panels:

Section 3.1.1:

Critical load exceedances for acidification for each of the four European (EU) models are shown in Figure 1 for 2010 and in Figure S3 (Supplement) for 2009, and Figure S9 (Supplement) for bias-corrected 2010. Figure 2 shows the reduced ensemble values for 2009 and 2010 (a,b), the bias-corrected value for 2010 (c), as well as common AQMEII4 domain total bar charts for all models and the reduced ensemble (d).

…

Bias correction for the reduced ensemble for the 2010 data resulted in substantial increases in predicted exceedances (compare last two columns of Figure 2(d), and compare Figure 1 to Figure S9). However, we note that the European data did not include speciated particulate matter and hence bias correction was not possible for part of the sulphur budget – much smaller impacts were noted for bias correction in North America where particulate sulphate data were available.

Section 3.1.2:

Critical load exceedances for eutrophication for each of the four EU models are shown in Figure 3 for 2010, in Figure S4 (Supplement) for 2009, and with bias-corrected deposition fields for 2010 in Figure S10 (Supplement). Figure 4 shows the reduced ensemble values for 2009 and 2010 (a,b), the bias-corrected values for 2010 (c), as well as common AQMEII4 domain summaries for all models and the ensembles (d).

…

The relative impact of bias correction was smaller than for acidification in terms of the total area in exceedance, but the magnitude of exceedances increased significantly (e.g. larger proportion of red to black areas in Figure 4(c) than Figure 4(b), and comparing the last two columns of Figure 4(d).

Section 3.1.3:

… and the domain summaries including bias corrected values for 2016 are shown in Figure 6.

…

The relative impact of bias correction was smaller than for acidification in terms of the total area in exceedance, but the magnitude of exceedances increased significantly (e.g. larger proportion of red to black areas in Figure 4(c) than Figure 4(b), comparing the last two columns of Figure 4(d), and comparing Figure 4 to Figure S10). Again, the higher levels of exceedance predicted for Europe may reflect the impact of the lack of particulate sulphate and particulate nitrate data for bias correction purposes….

Critical load exceedances with respect to the North American (NA) forest soil acidity for the years 2016 and 2010 are shown in Figures 5 and S5, respectively, the bias-corrected 2016 maps are in Figure S11, and the reduced ensemble maps for both years, and the domain summaries including bias corrected values for 2016, are shown in Figure 6.

…

The effect of bias correction was less pronounced than in Europe, and in general reduced the variability between model results. Note that unlike the European case, North American observation data used for bias correction included corrections for particulate sulphate air concentrations, allowing a greater degree of closure for the sulphur mass deposited. Comparing Figures 5 and S10 it can be seen that the bias correction has increased exceedances for the CMAQ and WRF-Chem simulations, and decreased exceedances for the GEM-MACH simulations, reducing the variability between the models. The extent to which model-to-model variability has been reduced subsequent to bias correction is also apparent in Figure 6(d) (bias correction exceedance bars are closer in size across models compared to pre-bias correction). The net result of bias correction being a slight increase in the area of exceedance in the reduced ensemble, comparing the two right-hand bars of Figure 6(d).

Section 3.1.4:

Exceedances with respect to the North American aquatic ecosystem CL dataset for the years 2016 and 2010 are shown in Figures 7 and S6, respectively, the bias-corrected maps for each model for 2016 are in Figure S12, and the reduced ensemble maps for both years and domain summaries including bias correction are shown in Figure 8.

…

The impact of bias correction on the North American aquatic ecosystems critical load exceedances was relatively minimal for the models included in the reduced ensemble: differences between Figures 7 and S12 are difficult to distinguish, and Figure 8(d) shows slight increases in the exceedances for CMAQ and WRF-Chem simulations, slight increases in GEM-MACH simulations, and a very small change in the reduced ensemble levels of exceedance.

Section 3.1.5:

Exceedances with respect to the USA CL of N for a 20% decline in sensitive epiphytic lichen species richness (221 eq-N ha^{-1} yr^{-1}) dataset for the years 2016 and 2010 are shown in Figures 9 and S7, respectively, bias-corrected 2016 values in Figure S13, and the reduced ensemble maps for both years and domain summaries included bias-corrected 2016 values are shown in Figure 10.

….

Bias correction values varied between the models, with CMAQ exceedances increasing slightly, GEM-MACH exceedances decreasing slightly, WRF-Chem exceedances increasing, and a slight increase in the overall extent and magnitude of the reduced ensemble exceedances in the last two columns of Figure 10(d). The similarity in the spatial distribution of exceedances is greater across models following bias correction (compare Figure 9 with Figure S13 (Supplement)).

Section 3.1.6:

Exceedances with respect to the USA CL of N for a decline in herbaceous species richness (436 to 1693 eq-N ha^{-1} yr^{-1}) dataset for the years 2016 and 2010 are shown in Figures 11 and S8, respectively, bias-corrected exceedances for 2016 appear in Figure S14 (Supplement), and the reduced ensemble maps for both years and domain summaries including bias correction for 2016 are shown in Figure 12.

…

The impacts of bias correction may be more easily distinguished for herbaceous species richness critical load exceedances compared to some of the other exceedance estimates (compare Figures 11 and S14), with the CMAQ and WRF-Chem exceedances increasing, and the GEM-MACH exceedances decreasing. The overall impact was a slight increase in the area and extent of the ensemble average exceedance (Figure 12(d)). “

A new section was also added to the Conclusions:

“Impact of Bias Correction as a Simple Form of Model-Measurement Fusion

A simple form of model-measurement fusion (bias correction) was applied to each of the models’ species contributing to total sulphur and nitrogen deposition, for those component species for which observations were available, and corresponding bias-corrected critical load estimates were generated. This sometimes resulted in substantial decreases in model-to-model variability in the CLEs generated, indicating that model-measurement fusion will decrease model-to-model variability, and improved CLE estimates, provided sufficient data is available on the main contributors to total sulphur and total nitrogen deposition. In the case of Europe, the application of bias-correction increased CLE variability for acidification, likely due to the lack of particulate sulphate observations in Europe for the years simulated. The substantial contrast to North American bias-corrected values suggests that the bias corrections for individual species contributing to total sulphur deposition may offset each other (e.g. positive biases in particle sulphate may be offset by negative biases in wet deposition). In the absence of speciated particle observation data in Europe, this compensating effect could not be captured using bias correction, and hence the European CLE variability increased with bias correction.

An important implication of the bias correction exercise conducted here is the need for observation data which close the sulphur and nitrogen deposition budgets to the greatest extent possible, when carrying out model-measurement fusion. The biases with respect to observations for sulphur species may reflect inaccuracies in the transformation of one species to another for example – if model-measurement fusion is applied to only some of the species contributing to sulphur deposition, the resulting total sulphur deposition field and exceedance estimates may be *less* accurate than the original model fields. Similarly, we note that the observations available here did not include particle nitrate or nitric acid data – and hence the impacts of model measurement fusion on total nitrogen deposition may potentially lead to *less* accurate estimates than the original model values. “

A small paragraph was added to the Abstract:

“Model-measurement fusion in the form of a simple bias correction was applied to the 2016 critical loads. This generally reduced variability between models. However, the bias correction exercise illustrated the need for observations which close the sulphur and nitrogen budgets in carrying out model-measurement fusion. Chemical transformations between different forms of sulphur and nitrogen in the atmosphere sometimes result in compensating biases in the resulting total sulphure and nitrogen deposition flux fields. If model-measurement fusion is only applied to some but not all of the fields contributing to total deposition of sulphur or nitrogen, the corrections may result in greater variability between models, or less accurate results for an ensemble of models, for those cases where an unobserved or unused observed component contributes significantly to predicted total deposition."

Some detailed comments are as follows:

Line 122-123: Deleting ", denitrification, nitrogen immobilization in the rooting zone, run-off volume, and a critical value of the non-sodium base cation to aluminum ion ratio"

The full line mentioned by Referee 2 here is "For example, in the steady-state or simple mass balance (SMB) model often used to define surface water critical loads for terrestrial ecosystems (Sverdrup and DeVries, 1994), BCdep includes the release of soil base cations due to weathering, non-marine chloride deposition, harvesting of base cation and/or nitrogen-containing biomass, denitrification, nitrogen immobilization in the rooting zone, run-off volume, and a critical value of the non-sodium base cation to aluminum ion ratio." That is, the Reviewer has asked for the removal of part of the *definition* of part of one of the terms used in CL calculations. While this is a standard definition, we are not able to remove parts of it – this would change the definition. Consequently, no change has been made to the manuscript in response to this comment.

Line 224: Delating ")"

Done (deleted).

Line 236: For the whole text, 10 in PM10 and 2.5 in PM2.5 should be in subscript.

Also done; both reviewers requested this, and this follows the convention used in other papers in ACP.

Line 272: Moving detailed introduction on critical load to the supplement or delete.

Done (see above general response).

Line 329, Figure 1: Adding explanation of the dashed lines.

Done. The added text, now in the Supplement, reads,

"Four Regions are displayed in the Figure. Region 1 corresponds to locations where nitrogen deposition has exceeded the $CL_{max}N$ value and sulphur deposition is always greater than $CL_{max}S$: the only means by which exceedances can be reduced is via reducing sulphur deposition to zero, and then nitrogen deposition to $CL_{max}N$.   In Region 2, a combination of non-zero reductions in sulphur and nitrogen deposition could be used to reduce exceedances. Region 3 exceedances can also be reduced by a combination of sulphur and nitrogen deposition reductions, though as the location of exceedance point E3 approaches the boundary with Region 4, more of the deposition reductions must come from sulphur deposition. In Region 4, reductions in nitrogen deposition will have no effect on exceedances; deposition reductions in sulphur must take place in order to prevent exceedances from occurring. The Regions thus denote different strategies that must be taken to prevent critical load exceedances. "

Line 681-700: The text can be shorter with Table 3.

Done. We've removed the text that summarizes and overlaps with the descriptions appearing in Table 3.

Line 892, Table 4: Can the emissions of major pollutants such as SO2, NOx and NH3 in each year be added? Also S and N deposition?

Note that this is with reference to revised manuscript Table 3. While noting that total S and total N deposition are given in later tables in the manuscript, we've repeated the summary totals and their range in the revised Table. Although we've mentioned that all models made use of the same starting inventories for emissions elsewhere in the document, we've repeated that information in the revised Table's caption.

Line 1067: Changing "workT" to "work."

Done.

Line 1103: Adding full point in the end.

Done.

Line 1170, Table 8: Moving this like tables to the supplement.

Table 8 contains the model performance metrics for speciated PM2.5 in North America, so we are assuming that this request is for us to move Table 6, Table 8, Table 9, Table 10, Table 12, and Table 14 to the Supplement. Done (renamed Tables S2, S3, S4, S5, S6, S7, respectively.

Line 1429: Missing ";" before ca.

I think the Referee means "," here, not ";". Modified with a ",".

Line 1438: Mission "." After "AQMEII4".

Done.

Line 1623: Deleting "HNO3 summer than".

Done.

Line 1779: The conclusion can be shorter.

The Conclusions were made more succinct (from 287 lines to 163 lines; a 43% reduction) by removing the sentences restating the numerical values of the exceedances, since those appear in the main body of the text, focusing on the most important points we wanted to make (removing detailed discussion of the reasoning behind those points, since those appear in the main text), and by modifying the recommendations at the end of the Conclusion section to only state the recommendations themselves, since the reasoning behind the recommendations also appear in the main text.

Line 1819, 1855: Here the underestimation of CLE caused by the bias of deposition modeling is of interest.

As noted above, bias correction was applied to determine the impacts of model bias on predicted CLE for both EU and NA in the years 2010 and 2016, and a few example model-measurement fusion references have also been added to the text. We thank the Referee for this suggestion: one key conclusion from the bias-correction exercise was that model-measurement fusion can reduce model-to-model variability in CLEs – with the key caveat that the all of the main contributors to S and N deposition need to be represented in the observations used for the fusion. This stood out in comparing the EU and NA bias corrected values, where the EU region lacked particle sulphate observations – and the EU bias correction resulted in greater model-to-model variability than the original model values

Line 1902, 1931: Same comment as above.

Same response as above.

Line 2636: The reference is repeated.

Corrected (removed 2nd, repeat reference).