Dear editor,

We hereby submit the final version of our manuscript. We thank you and the referees for the useful comments and suggestions made throughout the peer-review process.

Below, we provide point-by-point responses to your comments and those of the referees. A marked-up version of the manuscript is available at https://www.dropbox.com/s/n63wjuwjbv0hnaf/VanSundertetal_BG_MarkedUp.7z?dl=0.

Sincerely,

Kevin Van Sundert, on behalf of the co-authors.

OVERVIEW OF MAIN CHANGES

Following the suggestions made by the referees and the Associate Editor, we now substantially reduced the emphasis on the formerly defined goal to ultimately develop one single, global nutrient metric. Instead, in the Abstract we introduce our study as a first example ("manual") of how a nutrient availability metric may be evaluated and adjusted with inventory data. We also changed the wording to indicate that the metrics are adjustments of the IIASA-metric rather than improvements/updates per se, as the latter would require testing also for agricultural systems for which the IIASA-metric was initially developed. Overall, we agreed with most referee comments and we have made the necessary changes in the manuscript. In few cases where we did not agree (e.g. on the removal of gradients from the manuscript – COMMENTS 3.4 and 3.37), we thoroughly explain why.

POINT-BY-POINT REPLIES TO THE EDITOR AND REFEREE REVIEWS

COMMENTS RAISED BY THE ASSOCIATE EDITOR

COMMENT E.1: *"I agree with reviewer #1 that you should adjust your abstract. My recommendation follows the recommendation of this reviewer: to write a short convincing story that clarifies the reality of developing this kind of indices."*

We fully agree with the editor and reviewer and have rewritten the abstract accordingly. We first explain the need for nutrient availability metrics and discuss the research questions (i.e. detection of important soil variables, evaluation of the IIASA-metric and adjustment of the metric to the Swedish data), and then emphasize that our paper may serve as a manual for further evaluations and adjustments of nutrient metrics in the future (instead of stressing that we aim to develop one single, global metric (e.g. Lines 11-15 and 31-32)):

"Here, we use a Swedish forest inventory database that contains soil data and tree growth data for > 2500 forests across Sweden to (i) test which combination of soil factors best explains variation in tree growth, (ii) evaluate an existing metric of constraints on nutrient availability, and (iii) adjust this metric for boreal forest data. With (iii), we thus aimed to provide an adjustable nutrient metric, applicable for Sweden and with potential for elaboration to other regions."

"This study thus shows for the first time how nutrient availability metrics can be evaluated and adjusted for a particular ecosystem type, using a large-scale database."

As a side note, we want to mention here that development of a nutrient availability metric for forests not just limited to the boreal biome may not be so unrealistic. We are currently evaluating the metrics from the present paper against European ICP forest data, and preliminary analyses show that it is possible to further adjust our metrics such that they describe variation in normalized productivity across ICP data, without losing explanatory power for the Swedish inventory.

COMMENT E.2: *"Many details in your manuscript sometimes distract from the main story. I recommend to critically look if all the details are necessary."*

Based on this comment by the Associate Editor and remarks made by Ref#3, we removed repetitions and details, especially in the Discussion of the manuscript. We now for example only explain the normalization procedure once in the main text (COMMENT 3.17), removed unnecessary details in the discussion on soil C:N, and also shortened the discussion regarding research question 2 (i.e. evaluation of the IIASA-metric – COMMENT 3.2).

COMMENT E.3: *"Reviewer #3 does not like the term 'upgraded'. While I see the point made by this reviewer, I think this is mainly semantics. You may consider to use the term 'adjusted' or some other term, but if you give me a good reason why the term 'upgraded' is preferable, I am open for it."*

Following the suggestion made by Ref#3, we replaced the term "upgraded" by "adjusted" everywhere in the manuscript (in the text, captions and in Figs. 2, 10, 11, S5 and S6). Ref#3 states that "adjusted" is more appropriate than "upgraded", because our changes to the metric would merely represent a reparameterization, applied to a different land use type (i.e. Swedish boreal forest instead of agricultural field). Although we not only reparameterized, but also improved performance of the metric by including the soil C:N ratio, we understand Ref#3's comment and thus followed his/her advice.

COMMENTS RAISED BY REFEREE 1

COMMENT 1.1: "*Abstract - The abstract has not been sufficiently updated to accommodate changes in the revised manuscript. With the exception of the detail regarding the local fertility gradients, it reads largely like the original. Include information that describes some of the variables that were included/evaluated in the various metrics (other than SOC and C:N). Include mention of the value of stratifying based on N deposition and upland/wetland landscapes. Some discussion how the study sorted out co-varying factors such as N deposition and N-S climate would strengthen the study's utility to researchers working elsewhere with multiple covarying factors."*

Based on the comments by both referees and the Associate Editor, we rewrote parts of and added more information to the Abstract where necessary.

We now mention all variables included in the IIASA-metric (<u>Lines 21-23</u>):

"This IIASA-metric requires information on soil properties that are indicative of nutrient availability (SOC, soil texture, total exchangeable bases - TEB and pH) and is based on theoretical considerations that are also generally valid for non-agricultural ecosystems."

Moreover, before explaining the results in the Abstract, we now explicitly mention that methods for dealing with covarying climate, N deposition and soil oxygen availability are used in the manuscript (<u>Lines 15-16</u>). We however preferred not to explain the stratification and normalization procedures themselves in the abstract, because this would make us add too many details:

"While taking into account confounding factors such as climate, N deposition and soil oxygen availability, our analyses revealed that (…)"

In the final sentence, we now explain that our paper may serve as an example of how nutrient metrics can be evaluated and adjusted (Lines 31-32):

"This study thus shows for the first time how nutrient availability metrics can be evaluated and adjusted for a particular ecosystem type, using a large-scale database."

COMMENT 1.2: "*Line 10 - In my opinion, the contention that one metric is a goal remains rather over simplistic. Given the vastly different sorts of soils, climates, landscapes and production systems and management practices, a diversity of metrics seems more appropriate. I do agree that evaluation of the IIASA metric for forest systems in Northern Europe is worthwhile, but would suggest that the applicability of the study is more in showing how any particular metric is evaluated and modified, rather than how to develop a single metric."*

Development of one single, globally applicable metric is a great challenge, and may even be impossible. We therefore carefully changed our wording in the Abstract and Conclusion of the manuscript, and now refer to metrics applicable at large spatial scales (e.g. Lines 8-11 and 554-558):

"**Abstract.** The availability of nutrients is one of the factors that regulate terrestrial carbon cycling and modify ecosystem responses to environmental changes. Nonetheless, nutrient availability is often overlooked in climate-carbon cycle studies because it depends on the interplay of various soil factors that would ideally be comprised into metrics applicable at large spatial scales. Such metrics do currently not exist."

"The current nutrient availability metrics were developed based on data from Swedish conifer forests only, and can therefore not as such be extrapolated outside the boreal biome. In order to find out if development of a metric that compares the nutrient status across sites also beyond the boreal biome is feasible, the adjusted metrics developed in this study now need to be validated and further modified based on other forests for which the necessary soil information is available. In a later stage, this approach can then be expanded to other ecosystem types."

We also explain in the Abstract that our paper may serve as an example of how nutrient metrics can be evaluated and adjusted (Lines 31-32):

"This study thus shows for the first time how nutrient availability metrics can be evaluated and adjusted for a particular ecosystem type, using a large-scale database."

COMMENT 1.3: "*Line 16 - Define how forest productivity is measured (MAI m3/ha/yr??)."*

We added the units in the Abstract and made clear that mean annual volume increment was used as a proxy of forest productivity (Lines 17-18).

COMMENT 1.4: *"Line 19 - Define IIASA."*

Since the abbreviation "IIASA" is first mentioned in the Abstract, we now provide the full name there (Lines 20-21).

COMMENT 1.5: "*Line 21 - Insert 'forest' between normalized and productivity."*

We inserted the word "forest" as requested (Lines 18 and 24).

COMMENT 1.6: "*Line 33 - Remove 'for example'*"

Done. (Line 35)

COMMENT 1.7: "*Line 41 - Remove 'thus'*"

Done. (Line 43)

COMMENT 1.8: *"Line 45 - Replace 'have previously described' with 'commonly use'"*

Done. (Line 47)

COMMENT 1.9: *"Line 45 - Insert hyphen 'fertility-related'"*

Done. (Line 47)

COMMENT 1.10: "*Line 74 - Oxygen availability and relevance to wetland vs upland soils should be included in the uncertainty/applicability/future challenges."*

Done (Lines 513-517). We incorporated this in the paragraph on "Sources of uncertainty", since soil wetness and related oxygen availability increases variation in productivity, independent of nutrient availability (although we tried to account for this effect by performing stratified analyses):

"A more important source of uncertainty [than the number of replicates per data point] is probably the inevitable uncertainty related to the response variable, i.e. "climate-normalized" aboveground productivity. This includes uncertainty in the original productivity estimates (for which for example differences in management or disturbances likely increased variability) and additional variation caused by soil moisture effects on oxygen availability (which we accounted for by also performing analyses on split datasets). However, there is also uncertainty related to the normalization for climate: (…)"

COMMENT 1.11: *"Line 74-75 - Mention these [… environmental characteristics such as climate, rooting conditions and soil oxygen availability …] in the abstract"*

Done. Before explaining the results in the Abstract, we now explicitly mention that methods for dealing with covarying climate, N deposition and soil oxygen availability are used in the manuscript (Lines 15-16):

"While taking into account confounding factors such as climate, N deposition and soil oxygen availability, our analyses revealed that (…)"

COMMENT 1.12: "*Line 89 - Specify that it [the Swedish dataset] includes site information that allows comparison of the implications across soil oxygen/wetland, N deposition and climate gradients. Also integrate information regarding the local fertility gradients."*

We now specified this in the paragraph (Lines 89-95):

"Such a unique dataset – that comprises > 2500 conifer forest plots and thus provides sufficient statistical power for an evaluation of the metric – is provided by the Swedish forest inventory service. Moreover, it contains additional variables of interest related to N availability, such as soil total N stock and concentration, and especially the soil C:N ratio, which we expected to be an important factor in explaining variation in nutrient availability. This large dataset also allows evaluating our country-scale findings against local gradients in nutrient availability that avoid confounding effects of covarying factors such as climate and N deposition."

COMMENT 1.13: *“Line 148 - Insert: N deposition, soil landscape (wetland vs upland)”*

Done (Lines 151-153):

“Forest productivity across Sweden depends not only on soil nutrient availability, but also on climate, soil wetness and N deposition. Before evaluating the metric, we removed the influence of climate on forest productivity (“PRE” in Fig. 2). The influence of soil moisture and N deposition are considered in further analyses (see section 2.3.1).”

COMMENT 1.14: *“Line 274 - State the obvious parallels between the analysis of soil moisture class and the wetland soil types (histosol/podzol).”*

We now more explicitly show the parallels between soil moisture and type (Lines 288-294):

“Soil properties not only differed among soil moisture classes, but also among soil types. Especially histosols and podzols could be distinguished from the other soils: histosols (which largely overlapped with the wet soil moisture classes) were characterized by a low pH_{KCl}, high SOC and soil C:N ratio, while podzols were sandy and had a low TEB stock (Fig. S4). Differences in normalized productivity among soil types were observed as well. Histosols in particular showed reduced productivities compared to other soil types (Fig. 6). Hence, the wetness of a site and its type of soil (partly in parallel with wetness) could confound observed patterns in productivity associated with the soil variables and are therefore taken into account in the further analyses and their interpretation.”

COMMENT 1.15: *“Line 285 - Clearer less arbitrary wording would specify that “N deposition had a strong positive effect on normalized productivity based on method 2, but no significant effect with method 1.”*

We rephrased the sentences (Lines 299-303):

“However, N deposition did generally not have a significant effect on productivity normalized with method 1 (i.e. residual productivity), while with method 2 (i.e. actual/attainable productivity), there was a strong positive relationship with N deposition. The increasing N deposition along the north-south gradient in Sweden (e.g. Olsson et al., 2009) should thus be kept in mind when interpreting effects of soil variables on productivity when normalized following method 2.”

COMMENT 1.16: *“Line 297 - If the variables were not correlated, it does not matter what the trend was. This sentence should be removed.”*

Done. (Line 311)

COMMENT 1.17: *“Line 349 - Remove ‘y” in (Tables 4 and 5)”*

Done. (Line 365)

COMMENT 1.18: *“Line 385 - Good to connect soil moisture and soil type.”*

We appreciate the referee’s positive comment on this paragraph, where we link soil moisture and type.

COMMENT 1.19: *"Line 507 - Good place to mention the effects of variable forest management and age. Productivity relates to stand density (thinning operations) and stand age, and it would be good to acknowledge if/how those factors were addressed. The same goes for and more importantly for stand fertilization and soil drainage. Just review if this information was factored out (by site/data selection) or left in and part of unaccounted for variation."*

By taking the mean annual increment (MAI) over a rotation period as a proxy for aboveground productivity, the influence of stand age was largely removed. When we started the analyses for our paper, we formally tested if there was any remaining MAI-age relationship. This was not the case ($P > 0.05$).

Since we did not exclude any management type from the database, differences in management may explain part of the remaining variation. This information is included on Lines 513-517 of the revised manuscript:

"A more important source of uncertainty [than the number of replicates per data point] is probably the inevitable uncertainty related to the response variable, i.e. "climate-normalized" aboveground productivity. This includes uncertainty in the original productivity estimates (for which for example differences in management or disturbances likely increased variability) and additional variation caused by soil moisture effects on oxygen availability (which we accounted for by also performing analyses on split datasets). However, there is also uncertainty related to the normalization for climate: (…)"

Of the 23 Mha of productive forest in Sweden, only 44 000 ha was fertilized annually during 2006-2016 (https://www.slu.se/en/Collaborative-Centres-and-Projects/the-swedish-national-forest-inventory/forest-statistics/forest-statistics/). Stand fertilization is thus unlikely a major contributor to uncertainty in our database.

COMMENT 1.20: *"Fig. 4's Caption - Remove extra period after 'nitrogen deposition'."*

Done. (Fig. 4)

COMMENT 1.21: *"Fig. 4 - Check 'ln stock' between clay and TEB. Is that correct/complete? Add to caption?"*

Due to the alignment, it was unclear that ln(TEB stock) was meant here. We changed the alignment in Fig. 4b.

GENERAL COMMENTS RAISED BY REFEREE 3

COMMENT 3.1: *"The paper reads quite well, however, with some repetitions."*

Based on this remark and a comment by the Associate Editor, we removed some repetitions and details, especially in the Discussion. We now for example explain the normalization procedure only once in the main text (COMMENT 3.17), removed unnecessary details in the discussion on soil C:N, and also shortened the discussion regarding research question 2 (i.e. evaluation of the IIASA-metric – COMMENT 3.2).

COMMENT 3.2: *"Research question 2 seems to be a little bit overstressed as it should be clear from the dataset used, that the predominant part of the dataset is outside the range defined for the original IIASA-metric. This suggests, that the original IIASA-metric isn't applicable to the dataset. Thus, Q2 may be answered very shortly."*

We agree with the referee that the evaluation of the original IIASA-metric was not the main point of the study, and have therefore merged the two paragraphs on metric performance in the discussion into one (Lines 574-580):

"The IIASA-metric of constraints on nutrient availability does not clarify much variation in normalized productivity among Swedish forests. Moreover, SOC, soil texture, TEB and pH_w were apparently not optimally implemented. A low performance of the IIASA-metric in its current form for the Swedish database was expected, as it was initially developed for evaluating (constraints on) the soil fertility of agricultural ecosystems, and the Swedish database contains variable values outside the ranges to which the metric is sensitive. Soil conditions of agro-ecosystems indeed greatly differ from the boreal forests investigated in the present study. Many Swedish forest soils are for instance coarse-textured, and in addition, the database contains wet-soil forests, while arable soils are typically not water saturated."

COMMENT 3.3: "*In my view, Q3 should be regarded as a reparametrization or adjustment of the original IIASA-metric, not as an improvement."*

We agree with the referee and have replaced the term "upgraded" by "adjusted" throughout the manuscript.

COMMENT 3.4: "*I'm not very supportive of the analyses of the new metrics based on selected gradients as these – as far as I understood - contain no additional information, because they are sub-datasets of the complete dataset. Because they are – in my view – systematically (not randomly) selected, a somewhat higher degree of explained variance is not very surprising. Therefore, I suggest to omit this part of the manuscript."*

Based on an earlier remark by the Associate Editor, we decided to search for local nutrient availability/productivity gradients in the database, which offer the advantage that no (potentially R^2 reducing) normalization for climate would be necessary. If we would have chosen data points close to each other entirely randomly, R^2s would, as the referee suggests, probably be lower than was the case in the current analyses. However, in such case, the data points would together not represent a nutrient availability gradient anymore, and consequently variation in productivity would likely be dominated to a larger extent by "noise" such as uncertainty in productivity estimates, management, … We therefore did not remove the gradient analyses from the manuscript.

COMMENT 3.5: "*References are sufficient and up to date. However, more references dealing with forest issues would be an asset."*

We thank the referee for the positive evaluation of our use of references. Regarding citations on forest issues, we now included the studies by Laubhann et al. (2009) and de Vries et al. (2014) suggested by the referee, and in addition refer to more papers on boreal, European and global analyses of forest data, e.g.:

"Nutrients determine structure and functioning at all levels of biological organization. The availability of mineral elements influences plant growth (von Liebig, 1840), patterns of biodiversity (Fraser et al., 2015) and ecosystem processes (e.g. Janssens et al., 2010; Vicca et al., 2012; Fernández-Martínez et al., 2014). Moreover, nutrient availability can modify ecosystem responses to global atmospheric and climatic changes, such as nitrogen (N) deposition (**Nohrstedt, 2001; Hyvönen et al., 2008; Vadeboncoeur, 2010**), (…)" (Lines 34-38)

"Numerous studies have shown the strong influence of N deposition on forest productivity (e.g. **Laubhann et al., 2009; Solberg et al., 2009; de Vries et al., 2014**; Binkley and Högberg, 2016; **Wang et al., 2017**). Although N deposition can influence the soil properties considered in our analyses, it may also influence productivity without immediate changes in these soil properties (i.e. there is a time lag - Novotny et al., 2015). (…)" (Lines 192-195)

"Soil factors other than the soil C:N ratio and SOC either exhibited only a marginal influence on normalized productivity or their effect depended on the approach (Table 2). N stocks could explain variation across both methods, but their explanatory power was rather modest for method 1. We anticipate that if we aim to develop metrics applicable beyond the boreal biome, including N stock will be of limited value, as this variable is only loosely related to N availability (**Högberg et al., 2017**)." (Lines 446-449)

SPECIFIC COMMENTS RAISED BY REFEREE 3

COMMENT 3.6: *"Rows 24, 28, 91, 99, … - I wonder if 'upgraded' is the appropriate term here as it is more an adjustment from one land use type to another."*

We replaced the term "upgraded" by "adjusted" throughout the manuscript.

COMMENT 3.7: *"Row 35 - From et al. 2016 isn't explicitly a study on ecosystems responses to N deposition. It rather studies the effects of enhanced TN on productivity, which is later addressed in the manuscript. Probably, Laubhann et al. 2009 FORECO 258, de Vries et a. 2014 Curr Opinion Environ Sust 9-10 or related papers are more appropriate here."*

In response to this comment, we have replaced the citation to From et al. by Nohrstedt (2001), Hyvönen et al. (2008) and Vadeboncoeur (2010) to explain that ecosystem responses to N deposition depend on the background nutrient status (Lines 36-38). Even though these studies do not explicitly report on effects of N deposition, they refer to experimental N(PK) additions, which obviously show analogies with N deposition. These papers explain more clearly than From et al. (2016) that effects on tree growth, soil respiration and C sequestration are function of initial site fertility. To the best of our knowledge, the influence of background nutrient status on N deposition effects has never been assessed.

We now refer to Laubhann et al. (2009), Solberg et al. (2009), de Vries et al. (2014) and Wang et al. (2017) in the Methods (Lines 192-193) and Discussion (Lines 415-416) sections.

COMMENT 3.8: *"Row 70 - Abbreviations (TEB) need to be explained only at its first appearance."*

With this comment, the referee pointed to the line where both the full name and abbreviation of IIASA (not TEB) were mentioned for the first time in the Introduction. Since IIASA was already mentioned in the Abstract before, we removed the full name in Line 72. With the changes in our manuscript, TEB now appears for the first time in the Abstract too (Line 22), hence the full name + abbreviation is now given there.

COMMENT 3.9: *"Row 81-82 - Please explain, why the 'realism of the metric' is increased?"*

The IIASA-metric and our adjusted metrics give more weight to the soil variable with the lowest score. Following Liebig's Law of the Minimum, which states that it is the most limiting element that determines productivity, this way of weighing should improve realism of the metric. We now added an example why this is realistic to the paragraph (Lines 79-84):

"The species-specific score of the metric depends on four measurable soil variables, related to soil fertility: SOC (%), soil texture, TEB (cmol+ kg-1 dw) and pH measured in water (pHw). The metric score combines the scores of each of these four attributes (provided in a look-up table), but giving more weight to the attribute with the lowest score. Together with the non-linear relationships (e.g. for pH and SOC - see Methods), this increases the realism of the metric (cf. Liebig's Law of the Minimum (von Liebig, 1840); e.g. at optimal pH, the limiting effect of low SOC on plant growth will be stronger than in soils with very low or high pH where plant growth becomes more likely to be P-limited)."

In the context of differential nutrient limitation, we also provide another example with soil C:N in the discussion (Lines 539-541):

"This [giving more weight to the variable with the lowest score] corresponds to reality and enables accounting for the type of nutrient limitation. For instance, if soil C:N is high, indicating N limitation, the metric score will be substantially reduced by this high C:N, while at low C:N other limiting factors can dominate the metric score."

COMMENT 3.10: *"Row 89 - ... contains an additional variables ...' -> '... contains additional variables ..."*

Done. (Line 91)

COMMENT 3.11: *'Row 98 - ... in Sweden, ...' -> '... in Swedish forest soils, ...'*

The final part of the last introductory paragraph was edited and does not include the expression "in Sweden" anymore.

COMMENT 3.12: *"Row 114-117 - Why not use ERA-interim data for the determination of temperature sums (TSUM) to be consistent with precipitation data? As TSUM is determined here only from latitude, longitude and elevation using a parametrization obviously based on data prior to 1983, the consistency with other climate data is questionable."*

Prior to our further analyses such as normalization of productivity for climate etc., we tested which combination of climatological variables best described spatial variation in productivity. In the ERA-interim data, mean annual temperature (MAT) was available besides precipitation, but the regression model including TSUM instead of MAT performed better (i.e. it had a lower cross-validation based mean-squared error).

Odin et al. (1983) provide the original equation to calculate TSUM in Sweden based on latitude and elevation, but the actual equation given in our manuscript was based on a recent reparametrization. We now explicitly mention this in the text (Lines 113-116):

"To quantify the influence of climate on productivity across Sweden (question 1), we first determined the annual growing season temperature sum (TSUM) following a recently reparameterized version of the equation given in Odin et al. (1983), available on www.kunskapdirekt.se: (…)"

COMMENT 3.13: *"Row 119 - Please use consistent nomenclature throughout the manuscript (e.g. 'total nitrogen' or 'total N'; 'soil C.N ratio' or 'C:N ratio') and define abbreviations at its first appearance."*

Done. We now consistently refer to soil organic carbon concentration as SOC, to total N concentration as TN, to soil carbon to nitrogen ratio as soil C:N ratio etc.

COMMENT 3.14: "*Row 120 - Is the organic layer included in the depth intervals? Please clarify."*

Yes, the organic layer is included in the intervals. We now clarified this better in the text (Lines 122-124):

"(…) to values representative of the upper 10 cm (i.e. the 0–10 cm layer) and the upper 20 cm (i.e. the 0–20 cm layer) of the soil, including the organic layer."

COMMENT 3.15: "*Row 122 - Equation 3: Is 'humus stock' and 'humus depth' here synonymous to 'organic layer stock' and 'organic layer depth', respectively?"*

Yes. In order to be consistent with terminology across the manuscript, we renamed "humus" to "organic layer" in the equation (Line 125)

$$BD_{\text{organic horizon}}\ [\text{kg m}^{-3}] = \frac{organic\ layer\ stock\ [kg/m^2]}{organic\ layer\ depth\ [m]}, \qquad (3)$$

and in Table 1.

COMMENT 3.16: *"Row 152 - '... with ArcGIS (ESRI, 2011) …' seems not necessary here."*

We decided to keep the citation to correctly refer to the software used (Lines 157, 199 and 241).

COMMENT 3.17: *"Row 173-176 - This is a repetition of row 150 to 154 and row 168 to 169."*

The referee correctly points to information that was unnecessarily provided twice in the manuscript: the two methods for normalizing productivity were already introduced in Section 2.1 (General approach). In the following Section (Data analyses), we therefore shortened the paragraph to focus better on the main new point, i.e. that due to heteroscedasticity, we split the database by region for method 1, but not for method 2 (Lines 173-178):

"As explained in the paragraphs above, productivity was normalized using two methods. Method 1 considers the residuals to reflect deviations in productivity imposed by spatial variation in nutrient availability and in the absence of climate effects. However, residuals deviated more strongly from zero towards the warmer south (Fig. 3a), thus causing heteroscedasticity and a potential bias in the further analyses if not properly accounted for. For further analyses, we therefore split the database into three TSUM groups (north, middle and south; Fig. 3a). For method 2, considering the ratio actual/attainable productivity, this separation of different regions was not required."

COMMENT 3.18: "*Row 179 - '... (SOC concentration, total …' -> '… (SOC, total …'."*

As explained in COMMENT 3.13, we now consistently refer to soil organic carbon concentration as SOC, to total N concentration as TN, to soil carbon to nitrogen ratio as soil C:N etc.

COMMENT 3.19: “*Row 183 - ‘Soil type’ according to which classification?”*

Soil type was classified following the World Reference Base for Soil Resources. We now mention this in the paragraph (Line 190), in addition to the information added earlier under Table 1:

“We therefore tested if the selected soil variables and normalized productivity differed among soil moisture classes (…) and the most common World Reference Base for Soil Resources based soil types (histosols, gleysols, regosols, leptosols and podzols) …”

COMMENT 3.20: *“Row 186 - ‘Soil moisture classes’ according to which classification?”*

As specified under Table 1, soil moisture was determined in the field based on various indicators (e.g. groundwater depth, moisture at the surface, ground vegetation, elevated tree trunks, ...). The classification, specific for the Swedish inventories, is representative of the average moisture conditions during the growing season. We now mention that moisture classification is based on field indicators in the manuscript, and provide the references (Lines 187-189):

“We therefore tested if the selected soil variables and normalized productivity differed among soil moisture classes (dry, fresh, fresh-moist and moist, as available from the database and derived from a combination of indicators such as groundwater depth – Olsson, 1999; Olsson et al., 2009)”

COMMENT 3.21: *“Row 187-188 - Which classification was used here for soil type? If WRB is meant here (Table 1), the specifiers used (e.g. ‘peaty’, ‘wetland’) aren’t valid.”*

The referee correctly assumes that WRB was used for classification, as now explained in the text and under Table 1. We added this information to the paragraph (see COMMENT 3.19).

Although the specifiers we added to the soil types (peaty histosols, wetland gleysols, weakly developed regosols, gravel-rich leptosols and sandy podzols) are roughly correct, we agree that they were defined quite bluntly when WRB is used for classification. We therefore agreed with the referee’s request to remove the specifiers, but instead stressed the link between soil type and wetness better in the Results (Lines 288-294) and Discussion (Lines 406-414), as demanded by Ref#1:

“Soil properties not only differed among soil moisture classes, but also among soil types. Especially histosols and podzols could be distinguished from the other soils: histosols (which largely overlapped with the wet soil moisture classes) were characterized by a low pH_{KCl}, high SOC and soil C:N ratio, while podzols were sandy and had a low TEB stock (Fig. S4). Differences in normalized productivity among soil types were observed as well. Histosols in particular showed reduced productivities compared to other soil types (Fig. 6). Hence, the wetness of a site and its type of soil (partly in parallel with wetness) could confound observed patterns in productivity associated with the soil variables and are therefore taken into account in the further analyses and their interpretation.”

“In the same way as for soil moisture, stratification by soil type might help in resolving nutrient-productivity relationships. Soil properties and productivity differed among the five most common soil types in the database (i.e. histosols, gleysols, regosols, leptosols and podzols - Fig. S4). To some extent, these differences among soil types overlapped with these observed for soil moisture classes (e.g. wet histosols had the highest SOC, soil C:N and the lowest productivity), but additional patterns emerged as well (e.g. podzols had a particularly low TEB stock). Although actual differences in nutrient availability among soil types will in part underlie the variations in productivity, other factors related to soil type (e.g. wetness, soil depth or the rooting environment) may also influence productivity (Binkley and Hart, 1989). The main analyses of the current study were therefore stratified by both soil moisture and type to test the robustness of associations between nutrient related soil properties and normalized productivity.”

COMMENT 3.22: *"Row 179-180 - What is the rationale behind using total N concentration and total N stock as different variables, but only SOC concentration? First appearance of clay and sand fraction. Were sand and clay fractions derived from texture designation in the field or from lab analyses?"*

The idea of not only using TN (concentration), but also the amount of N per m^2 (stock) is that the latter may better represent the N present per unit soil volume. This matters for example when comparing SOM-rich soils with low bulk density with SOM-poor soils with high bulk density. In such case, higher TN might be observed for the former, while N stock is the same for both. For SOC, on the other hand, this is less of a problem, because soils with high SOC are roughly also the ones with the highest carbon stocks.

Originally, soil texture was present in the database as texture classes. In an earlier version of the database, however, these were approximately converted to percentages sand, silt and clay to facilitate analyses such as the ones performed in our study. We now mention this under Table 1 (Line 724):

"[c]In an earlier version of the database, percentages of sand, silt and clay were approximated from field based soil texture class."

COMMENT 3.23: "*Row 194-195 - Please amend the time period for which N deposition was available.*"

We used the most recent deposition data available for our analyses (i.e. from 2015). We added this information to our manuscript (Lines 196-199):

"To verify whether N deposition confounded our analyses, we extracted N deposition data of 2015 from a map available at http://www.smhi.se/sgn0102/miljoovervakning/kartvisare.php?lager=15DTOT_NOY___ (Swedish Meteorological and Hydrological Institute, 2018), using the ArcGIS software (ESRI, 2011)."

COMMENT 3.24: "*Row 204-206 - I don't understand the sentence (what is meant with 'each time'?). Please reformulate!*"

Like AIC(c), cross-validation can be used as a method to perform model selection. In the model selection procedure, we started with a model containing all explanatory variables, e.g.

$Y \sim a * X_1 + b * X_2 + c * X_3 + d * X_4 + e * X_5$ (model 1)

Then, the term with the lowest significance was removed, e.g. if this was c, then the model became:

$Y \sim a * X_1 + b * X_2 + d * X_4 + e * X_5$ (model 2)

To verify is model 2 was indeed better than model 1, we then compared the cross-validation based mean squared errors (mse) of the two models. If mse of model 2 < mse of model 1, we continued the procedure with model 2. With "each time", we thus meant that the selection procedure was carried out multiple times on different models, until the model with minimal mse was found. To make this clearer, we reformulated this in our manuscript (Lines 207-210):

"Starting from the full model containing all explanatory variables, the least significant term was removed, resulting in a simplified model. Performance of the full and simplified model was then compared using the mean squared error (mse), based on cross-validation (package DAAG - Maindonald and Braun, 2015). We repeated this model simplification procedure until mse stopped decreasing."

COMMENT 3.25: *"Row 233-245 - The selection of the 'local gradients' should be specified correctly. At row 235 and 239 it is claimed, that the local gradients were selected randomly, whereas at row 239 it is mentioned, that these were manually selected (for which it is hard to imagine that this is possible randomly). Moreover, at row 242 to 243, it is mentioned that 'we searched specifically for clear soil C:N gradients', again indicating a systematic selection procedure."*

We thank the reviewer for pointing out this apparent contradiction. We indeed manually (systematically) looked for gradients in soil moisture, TEB, productivity (and soil C:N) in ArcGIS, but within each "pool of potential gradients", we did not have a specific reason to prefer one gradient over another, which is why we wrote they were selected "randomly". In order to avoid confusion, we removed the word "random" in the revised manuscript, including Table and Figure captions (e.g. Lines 29, 237, 493, … and captions of Tables 4 and 5, Fig. S2).

COMMENT 3.26: *"Row 247-248 - Which 'additional tests' from which 'packages'?"*

This information was redundant and was therefore removed from the revised manuscript (the packages were explained in the following sentences - Lines 249-253):

"Moreover, for all regressions, potential non-linearities were detected with histograms of all variables' distributions and generalized additive models from the mgcv package (Wood, 2006). Data were log-transformed if their distribution was right-skewed, while polynomial (e.g. quadratic) functions were included in the model selection procedure where the general additive models suggested non-linear patterns. The variance inflation factor (package car - Fox and Weisberg, 2011) assessed possible multicollinearity."

COMMENT 3.27: *"Row 260-261 - 'TN' wasn't explained before."*

As explained in COMMENT 3.13, we now consistently refer to soil organic carbon concentration as SOC, to total N concentration as TN, to soil carbon to nitrogen ratio as soil C:N etc. We now provide the full name of TN at first appearance in the "Methods" section (Lines 121-124), and in the caption of (a.o.) Table 1, which introduces all variables:

"In order to facilitate between-site comparisons and to allow calculating the nutrient availability metric, we converted the soil measurements (SOC, soil texture, TEB, pH_w, pH_{KCl}, total nitrogen concentration (TN) and soil C:N ratio) taken per horizon to values representative of the upper 10 cm (i.e. the 0–10 cm layer) and the upper 20 cm (i.e. the 0–20 cm layer) of the soil, including the organic layer."

COMMENT 3.28: *"Row 262-263 - Is 'total N' the same as 'TN'?"*

Yes, we meant TN (total nitrogen concentration) with "total N" here. We replaced it by TN to be consistent across the manuscript.

COMMENT 3.29: *"Row 294 - Fig. 7b appears prior to Fig. 7a. Please change the numbering of the respective Figures accordingly."*

Fig. 7a/b represents the relationship between residual productivity and SOC/soil C:N. In an analogous way, Fig. 8a/b shows the relationship between actual/attainable productivity and SOC/soil C:N. In order to maintain this analogy between Figs. 7 and 8, we preferred to change the order of the sentences referring to SOC and soil C:N, rather than switching Figs 7a and 7b (Lines 307-310):

"For both SOC (Fig. 7a) and pH_{KCl}, the relationship with normalized productivity showed an optimum (i.e. an empirical quadratic relationship fitted better than a linear model). Normalized productivity was significantly negatively correlated with the soil C:N ratio (Fig. 7b), for which the effect became more pronounced towards the south (i.e. slopes and R²s increased; $F_{2,2274} = 34.23$, $P < 0.01$).”

COMMENT 3.30: “*Row 295-296 + 303 - The relationship between SOC (pH) and normalized productivity isn’t necessarily quadratic by nature (I guess, it only gives a better fit than linear). Please change wording. // Please use other wording for explanation of the functional type of empirical relations (i.e. ‘… was not quadratic…’).”*

We agree with the referee that the better fitting quadratic relationships of SOC and pH with normalized productivity are entirely empirical, and do not imply that the relationships are quadratic in reality. However, both show an optimum, and these optima do make sense for theoretical reasons discussed in the manuscript. As suggested by the referee, we rephrased earlier mentions to quadratic relationships and instead report that the variables showed an optimum effect on normalized productivity (e.g. Lines 307-308, 316-317 and 435):

“For both SOC (Fig. 7a) and pH_{KCl}, the relationship with normalized productivity showed an optimum (i.e. an empirical quadratic relationship fitted better than a linear model).”

“However, the function for pH_{KCl} did not show an optimum, but was linear with a significantly positive slope (Table 2).”

“The relationship of ln SOC with normalized productivity, which showed an optimum (Figs. 7a and 8a), (…)”

COMMENT 3.31: “*Row 303-305 - However, more than these two variables were included in the multiple regressions (Table 3).”*

We now made this clearer the manuscript (Lines 317-320):

“In summary, SOC and the soil C:N ratio were the only soil factors that showed a similar trend according both methods with an R^2 of at least a few percent, and were thus included in the multiple regression models for both methods 1 and 2 (these models also included other variables resulting from the stepwise regression analysis; Table 3).”

COMMENT 3.32: *“Row 349 - Is Table 5 meant here instead of Y5?”*

Done. (Line 365)

COMMENT 3.33: *“Row 383-384 - the specifiers of the soil type (e.g. peaty, wetland) don’t appear in Fig. S4.”*

In accordance with the referee’s earlier COMMENT 3.21, we removed the specifiers from the main text.

COMMENT 3.34: *“Row 395-398 - However, as N deposition is affecting nutrient availability (which is demonstrated by the positive correlation to SOC and TN – Fig. 4), also the effect of – in this case – SOC may be partly removed by the normalization procedure (method 1).”*

We agree with the referee that some minor influence of SOC on productivity may be removed by normalizing, because by removing climate influence following method 1, also N deposition influence is largely removed, and SOC correlates with N deposition (Pearson’s r = 0.30). However, there is considerable variation in SOC irrespective of N deposition, such that the main effect of SOC on productivity nevertheless remains.

COMMENT 3.35: *"Row 466-474 - The 'not optimal implementation' of SOC (and other variables) is not very surprising as the database contains a substantial part outside the range of the original IIASA-metric. Extrapolation beyond the range of empirical functions should usually be avoided."*

IIASA does not explicitly provide information on what soil variable ranges the metric can be applied, but we agree with the referee that data that fall outside of the original metric's ranges explain in part why the variables seem suboptimally implemented. We therefore added this to the paragraph (Lines 474-480):

"The IIASA-metric of constraints on nutrient availability does not clarify much variation in normalized productivity among Swedish forests. Moreover, SOC, soil texture, TEB and pHw were apparently not optimally implemented. A low performance of the IIASA-metric in its current form for the Swedish database was expected, as it was initially developed for evaluating (constraints on) the soil fertility of agricultural ecosystems, and the Swedish database contains variable values outside the ranges to which the metric is sensitive. Soil conditions of agro-ecosystems indeed greatly differ from the boreal forests investigated in the present study. Many Swedish forest soils are for instance coarse-textured, and in addition, the database contains wet-soil forests, while arable soils are typically not water saturated."

COMMENT 3.36: *"Row 476-481 - Basically, this is a reparametrization of the original IIASA-metric."*

We agree with the referee that the altered equations for SOC and pH are actually reparametrizations. However, we also included the soil C:N ratio, which is new to the metric and which, as we demonstrate, is an important indicator of nutrient availability for forests in Sweden. Based on this and other comments of the referee, we have reworded this sentence to indicate that the metric was "adjusted" instead of "improved" (Lines 482-484):

"Based on results of the analyses for question 1, the nutrient availability metric was adjusted by i) including an empirical optimum in the influence of SOC on normalized productivity, and ii) including soil C:N, thus more explicitly incorporating the availability of N."

COMMENT 3.37: *"Row 488-489 + 489-491 – This [the fact that the adjusted metrics described variation in productivity for gradients] is again not surprising as the gradients are sub-datasets of the complete dataset. // I fear, the higher R^2s are most likely because they were not randomly selected. If they were randomly selected, I would guess that the probability of higher or lower R^2s is almost similar."*

Based on an earlier remark by the Associate Editor, we decided to search for local nutrient availability/productivity gradients in the database, which offer the advantage that no (potentially R^2 reducing) normalization for climate would be necessary. If we would have chosen data points close to each other entirely randomly, R^2s would, as the referee suggests, probably be lower than was the case in the current analyses. However, in such case, the data points would together not represent a nutrient availability gradient anymore, and consequently variation in productivity would likely be dominated to a larger extent by "noise" such as uncertainty in productivity estimates, management, … We therefore did not remove the gradient analyses from the manuscript.

COMMENT 3.38: *"Row 496-498 - As far as I understood, both approaches are entirely empirical. Moreover, I would expect that a multiple regression approach can be updated to datasets from other ecosystems as well."*

We agree with the referee that the adjusted metrics are definitely not mechanistic models, but rather a weighted set of empirical regression equations. Reparametrizing the metrics however remains more obvious than doing the same for multiple regressions, especially with regards to differential nutrient limitation, as also mentioned under COMMENT 3.9 and in the Discussion (Lines 539-541):

"This [giving more weight to the variable with the lowest score] corresponds to reality and enables accounting for the type of nutrient limitation. For instance, if soil C:N is high, indicating N limitation, the metric score will be substantially reduced by this high C:N, while at low C:N other limiting factors can dominate the metric score."

With multiple regression equations, the underlying model structure (including non-linear functions and interactions) would likely completely shift depending on the type of nutrient limitation. Consequently, this would require more than a simple reparametrization or addition of a variable not yet included. To clarify the need for the more complicated structure compared to regressions, we have included a few sentences on this in the discussion (Lines 498-506):

"Variation in normalized productivity explained by the adjusted metrics (R^2 = 0.03–0.21 and R^2 = 0.06–0.18) was similar to the variation explained by multiple regression equations (R^2 = 0.18–0.22) that contained the same (and more) soil variables than the metrics. The metrics, however, have the advantage that they can be updated more easily than equations from multiple regressions, especially if additional soil parameters need to be included for other ecosystems. Moreover, the interaction effect – with the highest weight for the least optimal soil parameter – cannot be mimicked with a multiple regression approach. In order to further adjust the metrics, and to test to what extent they can already describe variation in nutrient availability outside of Swedish conifer forests, additional datasets with productivity and soil information are needed. Such datasets include large-scale inventories such as the one considered in the present study, but also local gradients and nutrient manipulation experiments. The latter two have lower generalizability, but offer the advantage that normalization for climate is not needed."

COMMENT 3.39: *"Row 518-520 - What are the main sources of uncertainty in the attainable productivity estimates?"*

Attainable productivity was extracted from a map, provided by Bergh et al. (2005). The authors estimated attainable productivity for Norway spruce in Sweden by combining climate/environmental information with an experimentally derived relationship between productivity and intercepted radiation. This relationship was based on only a few nutrient optimization experiments. Hence, large uncertainty can be expected in the national generalization. We shortly added this information to the manuscript (Lines 517-526):

"However, there is also uncertainty related to the normalization for climate: by taking residuals of the productivity vs climate regression model (method 1), we for instance unintentionally not only removed the direct effect of climate on productivity, but also its indirect effect through nutrient availability. Normalized productivity based on this method thus mainly represents productivity as influenced by regional variation in nutrient availability. The approach taking actual/attainable productivity as a response variable (method 2) does not suffer from this issue, but there the estimates of attainable productivity come with a high uncertainty, as they were based on only limited experimental data to establish a relationship between productivity and intercepted radiation. As a consequence, the low R^2 values are partly due to shortcomings of the normalization procedure that can only be overcome by using datasets where climate does not vary but nutrient availability does. Such datasets are provided by local gradients, such as the five local nutrient availability gradients that we selected from our database for additional evaluation of our adjusted metrics."

COMMENT 3.40: *"Figure 2 - '... vs climate and species; ...'-> '... vs climate and species (spp.); ...'."*

Done.