Dear editor,

We hereby submit the revised version of our manuscript. We are grateful to you and both referees for the insightful comments that helped us to significantly improve our manuscript.

Below, we provide point-by-point responses to all General, Specific and Technical Comments. A marked-up version of the manuscript is available at https://www.dropbox.com/s/kf5qqqkxrau6ofq/KevinVanSundert_etal_Biogeosciences_MarkedUp.7z?dl=0.

Sincerely,

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

OVERVIEW OF MAIN CHANGES

Based on several excellent suggestions of the referees and the editor, the revised manuscript now includes new analyses in which the influence of soil type, soil moisture and nitrogen deposition is thoroughly explored. We added a section 'Identifying potentially confounding factors', in which we investigate and discuss (confounding) factors that should be taken into account when evaluating nutrient-productivity relationships. Referee 1 and the editor for example advised to stratify by soil moisture and soil type. In the new section, we first present how soil properties and normalized productivity are related to soil moisture and soil type (Sections 2.3.1, 3.1 and 4.1). In the next sections we then present and interpret the stratified analyses in view of our research questions (e.g. Lines 306-308, 337-340, 487-488, 524-526). Referee 1 and the editor also emphasized that the role of N deposition should be discussed. We therefore investigated how N deposition correlated with soil variables and (normalized) productivity (e.g. Lines 189-198, 280-289, 391-402). This revealed a strong positive correlation between N deposition and productivity. However, this correlation disappeared when productivity was normalized for climate following method 1 (i.e. "residual productivity" - Fig. 3a), while it remained for the normalization according to method 2 (i.e. actual/attainable productivity), since actual/attainable productivity increases from north to south in Sweden, together with N deposition. In other words, confounding N deposition effects on nutrient-productivity relationships were neutralized for method 1, whereas for method 2, confounding N deposition should be kept in mind when evaluating nutrient-productivity relationships.

In the previous version of our manuscript and in our replies, we had mentioned the future use of data from natural gradients in nutrient availability. The editor wondered why we did not already evaluate our metrics[1] against data from gradients within the Swedish data. We thank the editor for this excellent idea, and have now evaluated our upgraded metrics against data from five nutrient availability/productivity gradients in Sweden. These gradients were selected to cover a range in nutrient availability and productivity within the same/very similar climate. The results from these analyses indicated that tree productivity was significantly related to our metrics for all gradients (Tables 4 and 5). Moreover, the R^2s of these relationships were generally higher than was the case when considering the complete database (R^2s were up to 38

[1]We now present two upgraded nutrient metrics instead of one, following referee 2's suggestion (COMMENT 2.16) to develop a second metric using productivity normalized according to method 2 (actual/attainable productivity - Fig. 3b), apart from the other metric (using productivity normalized following method 1, i.e. residual productivity - Fig. 3a) already presented in the previous version of the manuscript.

%). This suggests that the (necessary) procedure to normalize productivity for factors like climate in the previous analyses reduced the predictive power of the metrics. This last aspect also relates to referee 2's comment on the low R^2's, which we have now addressed in a section on sources of uncertainty and future challenges in the discussion (Section 4.5).

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

GENERAL COMMENTS RAISED BY REFEREE 1

COMMENT 1.1: *"(...) Owing to the strong influence of site wetness on SOC, it is advisable to stratify and analyze soil fertility x productivity relations by wetness class, not just along N-S gradient."*

We followed the referee's advice to stratify the main analyses by soil moisture, and we have also stratified by soil type (separately), as suggested by the editor. In the revised manuscript, we present the stratified regression analyses (on soil variables + N deposition) and evaluations of the upgraded metrics. The regression outputs (Tables S3-S6 and S12-S15) generally confirmed the robustness of our results. We briefly discuss outcomes of the stratified analyses in the manuscript (e.g. Lines 306-308, 337-340, 487-488, 524-526).

COMMENT 1.2: *"N inputs are 4-5 times higher in Southern than Northern Sweden (Binkley and Hogberg 2016) and thus present a further factor that should be considered with evaluating nutrient and forest productivity across these sites (de Vries et al. 2014)."*

We tested the influence of N deposition on productivity, and on climate-normalized productivity. As explained on Lines 280-289 in the manuscript, N deposition indeed showed a strong positive relationship with productivity, because productivity increases in the north-south direction, as do N deposition, temperature and incident light. Whether N deposition correlated with normalized productivity depended on the normalization procedure (Fig. 3): for method 1 (which normalized productivity for species and climate [+ indirectly N deposition!]), there were mostly no positive associations with N deposition, while for method 2 (actual/attainable productivity of spruce), there was a strong positive relationship. This positive relationship followed logically from the increase of the ratio actual/attainable productivity from north to south Sweden. These issues, and their consequences, are discussed in the revised manuscript on Lines 391-402:

"Many studies have shown the strong influence of N deposition on forest productivity (e.g. Binkley and Hogberg, 2016; From et al., 2016). As expected, N deposition correlated to some extent with some of the soil variables considered in the present study, such as the total soil N stock and concentration (Fig. 4b). Furthermore, N deposition was strongly positively related to productivity. However, this effect of N deposition on productivity cannot be separated from the influence of climate and light, as all these factors increase together in the north-south direction. Nevertheless, we argue that for the goals of this study, i.e. investigating soil nutrient-productivity relationships across Sweden and developing a nutrient metric, the spatially varying N deposition is not problematic, since the normalization for climate and species according to method 1 (Fig. 3a) at the same time also removed the influence of the confounding N deposition on productivity. Accordingly, "Residual productivity" was generally not correlated with N deposition (Table S3). The response variable derived from method 2 (i.e. actual/attainable productivities for spruce - Fig. 3b) in contrast, correlated strongly with N deposition (Table S4), because both actual/attainable productivity and N deposition increased from north to south. Consequently, relationships between actual/attainable productivity and soil data for this method were unavoidably confounded by N deposition."

COMMENT 1.3: *"(…) Based on the conditions under which IIASA was developed, how well might it be expected to perform on forest soils and with tree growth? Without clear description of the IIASA (currently some found in Intro, but lacking from Abstract), the paper presents a 'house of cards' based on a relatively unknown, and possibly weak foundation. Finally, please justify why the simpler approach of originating with 'no' tool and creating one from the ground up with existing data, as outlined in Question 1, is not both adequate and preferable."*

We agree with the referee that this choice required further clarification. In the revised manuscript, we therefore explain the basic principle of the IIASA-metric in more detail in the abstract (Lines 17-21), methods (Lines 130-146) and discussion (Lines 446-451):

"For the metric, we started from a (yet unvalidated) metric for constraints on nutrient availability that was previously developed by IIASA (Laxenburg, Austria). This IIASA-metric - initially developed for evaluating potential productivity of arable land - consists of soil properties that are indicative of nutrient availability and is based on theoretical considerations that are also generally valid for non-agricultural ecosystems."

"The IIASA-metric of constraints on nutrient availability, originally meant for use on arable land, incorporates four crop specific scores (estimated for SOC, texture, TEB and pHw) that can be assigned to a soil (IIASA and FAO, 2012). These scores, which can be found in look-up tables (http://webarchive.iiasa.ac.at/Research/LUC/GAEZv3.0/soil_evaluation.html), were derived from crop growth data on different agricultural soils. Given that we were analyzing boreal forests and not crops, we averaged the scores of the different crop species for each of the four soil properties. As such, we thus removed crop specific requirements, but generally known relationships between the soil variables and plant performance (not only valid for agro-ecosystems), such as an optimum for pH, remained. (…)"

"(…) the structures of its formulas (Eqs.(6-9)) reflect general mechanisms that link soil properties to nutrient availability, which are also valid for non-agricultural ecosystems. Soil pH for example shows a typical optimum effect on nutrient availability, while SOC and TEB have a direct positive non-linear influence (IIASA and FAO, 2012). The final weighing of the four partial scores (Eq. (10)) finds its rationale in the idea that if a certain soil property is particularly suboptimal, it will be the most important nutrient-related determinant of productivity, with less influence of the other soil properties that are within the optimal range. (…)"

and added paragraphs in the discussion on why (i) we started with this particular metric (formulas were based on theoretical considerations, it indirectly includes interactions, is upgradeable, and it is the only generic nutrient metric - Lines 444-455):

"Although the IIASA-metric of constraints of nutrient availability was originally designed for evaluating constraints on nutrient availability of arable lands, we opted to start with this metric for a couple of reasons. Apart from the fact that to our knowledge, it represents only attempt so far to develop a generic nutrient metric, the structures of its formulas (Eqs.(6-9)) reflect general mechanisms that link soil properties to nutrient availability, which are also valid for non-agricultural ecosystems. Soil pH for example shows a typical optimum effect on nutrient availability, while SOC and TEB have a direct positive non-linear influence (IIASA and FAO, 2012). The final weighing of the four partial scores (Eq. (10)) finds its rationale in the idea that if a certain soil property is particularly suboptimal, it will be the most important nutrient-related determinant of productivity, with less influence of the other soil properties that are within the optimal range. This way of weighing can be considered as a type of interaction, but one that cannot be implemented in a simple linear regression model. Hence, our main reason for adopting the IIASA-metric as a starting point is that, in spite of its simplicity, it is based on theoretical considerations. Moreover, adopting this structure allows for updating, in contrast to multiple regression equations (see below)."

and why (ii) multiple regression equations are not practical (i.e. not upgradeable with data from outside Sweden – Lines 494-505):

"Variation in normalized productivity explained by the upgraded 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 can, however, easily be updated, while equations derived from multiple regressions would be of less use for building a nutrient availability metric because such an entirely empirical approach does not allow for later updates of parameters or model structures based on data from other ecosystems. In order to further enhance performance of the metrics, and to test to what extent they can describe variation in nutrient availability outside of Swedish conifer forests, additional datasets with productivity and soil information are needed. (…)"

SPECIFIC AND TECHNICAL COMMENTS RAISED BY REFEREE 1

COMMENT 1.4: *"Line 8 - Nutrients are one of the factors that influence C cycling. As presented this suggests they are the only or the primary factor. (...)"*

We rephrased the sentence (<u>Lines 8-9</u>).

COMMENT 1.5: "*Line 10 - 'ideally' The potential value of simplifying ecosystem complexity into interpretable and useful metrics is critically important. However, given the complexity of soils and forest ecosystems, it is worth questioning whether a single fertility metric is an attainable goal. Rather than assuming such an index is 'ideal' this paper might sim-ply evaluate relationships between soil fertility properties and forest growth are robust and broadly useful."*

We were glad to see that the referee appreciates our effort to develop a common nutrient availability metric. Whether or not a global, generic nutrient availability metric can be developed remains an open question. In any case, even if this turns out too complicated, a valuable alternative could be to develop a metric for forests only for example. In response to this comment, we have specified some challenges in the development (e.g. differential nutrient limitation and low R^2*s* resulting from the normalization of productivity) in <u>Section 4.5</u> in the revised manuscript.

Moreover, although the development of a nutrient metric is a key part of our study, our manuscript also provides important information about factors related to soil fertility in general. In the paper, we demonstrated for example that the soil variables most critical for predicting productivity across Swedish forests were soil C:N ratio and SOC.

COMMENT 1.6: *"Line 10 - 'Such a metric does not exist' Is the lack of such a metric a major obstacle to understanding of forest production? An objective of this paper should be to quantify how such a metric improves our ability to predict forest growth."*

The need for a common metric of nutrient availability goes far beyond predicting forest productivity. A generic nutrient availability metric would allow to make comparisons in the nutrient status among monitoring sites and field experiments. This would for instance greatly increase the statistical power of synthesis studies that aim to elucidate the role of nutrients in ecosystem functioning (e.g. <u>Lines 32-50</u>):

"Nutrients determine structure and functioning at all levels of biological organization. The availability of mineral elements influences for example 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 (From et al., 2016), increasing CO_2 levels (Norby et al., 2010; Terrer et al., 2016), warming (Dieleman et al., 2012) and drought (Friedrich et al., 2012). Given the crucial role of nutrients in terrestrial carbon cycling and in shaping the magnitude and direction of its feedbacks to climate change, nutrient availability should be taken into account in global analyses and in Earth system models (Goll et al., 2012; Thomas et al., 2015; Wieder et al., 2015). This is, however, not yet common practice because we often lack the soil data and metrics needed to accurately account for nutrient availability.

Comparing nutrient availability among terrestrial ecosystems is thus difficult for two reasons: comprehensive and harmonized data on soil properties and nutrients are not usually available from experimental and observational sites, and no standardized quantitative metric exists to compare the nutrient statuses of terrestrial ecosystems at the global scale, or even at a national scale (e.g. for Sweden, which is considered in the present paper). In the absence of a standardized nutrient availability metric, studies comparing nutrient availability across sites have previously described soil fertility related approximations such as the height of 100 year old trees (which, however, also depends on other factors such as soil depth and hydrology - Hägglund and Lundmark, 1977) or have manually classified sites as low, medium, and high nutrient availability based on existing site information (Vicca et al., 2012; Fernández-Martínez et al., 2014). The absence of a more nuanced expression impedes elucidating the role of nutrient availability in ecosystem processes and functioning (Cleveland et al., 2011) and how these respond to global change, and precludes investigating non-linear effects of nutrient availability."

COMMENT 1.7: *"Line 12 - Insert 'of' between 'combination' and 'soil' Change 'plant' to 'tree' Insert 'forest' before 'nutrient'"*

Done. (Lines 11-13)

COMMENT 1.8: "*Line 13-14 - Developing this analysis using an unproven metric is counterin-tuitive. Justify why you did not start out with 'no' tool and build one from the ground up with existing data. Also, explain how the IIASA approach tool developed elsewhere for crops has application for Swedish forests. Clarify how advances from your analysis are independent of potential weakness or limitations of the IIASA approach.*"

In the revised manuscript, we now better explain why we opted to start with the IIASA-metric: the metric's scores for soil properties are based on generally (qualitatively) known relationships between soil properties and soil fertility, which are to some extent valid across ecosystem types (Lines 17-21, 130-142, 444-455). We now also clarified that we anticipated in advance that N availability (~ soil C:N ratio) should be better represented in the metric (Lines 459-465). We further stress that the metrics are upgradeable - in contrast to multiple regression equations, see Lines 494-505 - and can be tested against data from outside Sweden. Also see COMMENT 1.3.

COMMENT 1.9: *"Line 19 - 'not well implemented' Define what that means, especially in light of whether or IIASA provides a useful platform to work with."*

If soil variables are properly implemented in a nutrient metric, they should not be correlated with residuals obtained by regressions of (normalized) productivity vs that metric. We clarified this in the manuscript (Lines 22-23, 161-163, 213-216, 312-318, 340-346, 349-351, 466-469) and in the captions of Tables S9 and S16-19.

COMMENT 1.10: *"Line 37/38 - Site index is typically an estimate of site productivity, rather than soil fertility per se. The terms are certainly related, but key differences might include factors such as soil depth and hydrology."*

We agree and thank the reviewer for notifying. We rephrased the sentence (Lines 45-47).

COMMENT 1.11: *"Line 47 - 'are more indicative' it's not clear of what. Please revise, clarify."*

Soil properties such as soil texture, SOM and pH are more indicative of the general nutrient status, because nutrient availability is determined by the interplay between different soil nutrients and influenced by soil properties like pH. We rephrased this in the manuscript (Lines 54-57).

COMMENT 1.12: *"Line 100 - What was the size of the forest plots? Are these all plantations or are some natural stands? How are management interventions such as fertilization and thinning addressed and controlled for? What is the average and range of stand age and stand density or basal area?"*

We added this information to Table 1 in the "Methods" section. As the inventory represents a random sample of all Swedish forests, the majority is managed. There are typically 1-2 thinnings per rotation, while fertilization was only carried out for a minor proportion of the forests. The thinnings did in all likelihood not bias our results, because we used rotation-averaged productivities, based on height development curves of dominant trees, as our response variable.

COMMENT 1.13: *"Line 122 - Before suggesting that metric scores can validly 'be assigned to any soil' it would be useful to know more about the data that went into development of the IIASA and how well those soil types and crop conditions match the forest soils of Sweden."*

We fully agree with the referee. In the revised manuscript, we now explain that the IIASA-metric's formulas represent general relationships between productivity and a few soil properties (Lines 17-21, 130-142, 444-455). These equations were calibrated based on data from different agricultural soils and crop growth. Although forest soils differ in various aspects (e.g. Lines 461-465), general relationships between SOC, soil texture, TEB, and pH on the one hand and productivity on the other hand should in principle remain qualitatively the same, unless other confounding factors (such as soil moisture) mask these relationships.

COMMENT 1.14: *"Line 123 - This suggests that the IIASA metric is not sensitive to crop specific requirements."*

We rephrased this (Lines 133-137). The IIASA-metric is crop-specific, but general, logical patterns between soil properties and the partial scores remain after averaging scores of different species.

COMMENT 1.15: *"Line 284-285 - Use soil moisture classes to stratify fertility by productivity. There appear to be abrupt differences between upland and 'wetland' forests that are at least as significant as the N-S gradients."*

We thank the referee for this insightful comment. We have incorporated the suggested analyses in the manuscript, and in addition stratified by soil type (as suggested by the editor).

We added a section 'Identifying potentially confounding factors', in which we investigate and discuss (confounding) factors that should be taken into account when evaluating nutrient-productivity relationships. In the new section, we first present how soil properties and normalized productivity are related to soil moisture and soil type (Sections 2.3.1, 3.1 and 4.1). In the next sections we then present and interpret the stratified analyses in view of our research questions (e.g. Lines 306-308, 337-340, 487-488, 524-526). In general, the stratified analyses confirmed the robustness of our results.

COMMENT 1.16: *"Line 295 - The high SOC of wet soils is less a limitation than the anaerobic rooting environment."*

We agree with the referee and have rephrased the sentence accordingly (Lines 421-424).

COMMENT 1.17: *"Line 296 - Consider optima for soil saturation, soil oxygen, or soil moisture like the one currently used for pH. This is independent of SOC."*

We agree with the referee that if we were to develop a productivity index, including soil moisture or oxygen should be considered. Our goal, however, is to develop a metric on nutrient availability. Distinguishing nutrient availability (metrics) from soil moisture (metrics) is essential if the eventual metric is to be used to disentangle drivers of ecosystem functioning and to separate and quantify the influence of nutrient availability and soil moisture. Although soil saturation/moisture/oxygen influence productivity, we therefore did not include these factors in the upgraded metrics.

COMMENT 1.18: *"Line 301 - It is ironic that the only nutrient examined is of limited value. It should be stressed that total N is a much larger pool than the N that's readily available to plants. The complications of inorganic N inputs across the N deposition gradient, and inorganic N losses (leaching and denitrification) across wetness/saturation gradient are likely considerable."*

We fully agree with the referee and have rephrased the sentence to make clear that available N is just a small fraction of total N (Lines 429-431).

COMMENT 1.19: *"Line 304 - This relation appears to stem on the large difference in clay in Wet spruce sites (Fig S3e). Again, stratifying wetland from up-land forests might isolate and better resolve these patterns."*

It is not entirely clear to us what the referee refers to here: there was a significant positive relationship between the mineral soil clay fraction and normalized productivity (albeit with very low R^2s). The referee correctly observed that mineral soils at wet spruce sites had the highest clay contents, but based on this and the fact that productivity was suppressed at wet sites, we would rather expect a negative association between mineral soil clay and normalized productivity.

Following the referee's advice, we stratified analyses by soil moisture. The positive relationship between mineral soil clay and normalized productivity remained, irrespective of the moisture class, or became non-significant (Table S5).

COMMENT 1.20: *"Line 309 - Insert and stress "in the absence of direct soil nutrient data"*

We followed the referees suggestion and have rephrased the sentence (Line 438).

COMMENT 1.21: *"Line 315/316 - General differences between arable and Swedish forest soils could be used to define expectations regarding the utility of the IIASA method and probable modifications. Saturated soils, low-clay soils, low N soils might all define some of those differences."*

We thank the referee for this suggestion and added these differences between arable and conifer forest soils to the discussion about the performance of the IIASA-metric (Lines 462-465).

COMMENT 1.22: *"Line 324 - High water table, low soil oxygen, soil saturation, wetland soils."*

We clarified this link now at various places in the manuscript (Lines 184-185, 372-374, 417-420, 470-473, 543-544).

COMMENT 1.23: *"Table 1 - Units are included for everything but TN, and moisture. How was moisture determined and how well does it estimate relevant seasonal moisture dynamics?"*

We thank the referee for noticing this and added the unit for TN [%]. Soil moisture and soil type do not have a unit, but are represented in different classes. This is now indicated in Table 1. We provided relevant information about the plots, sampling, lab analyses and classification in the caption of the table, and in the notes underneath it. The classification of soil wetness is representative of the average moisture conditions during the growing season.

COMMENT 1.24: *"Table 2 - Sample depths are listed for everything but sand and clay."*

Soil texture data were available as texture classes for the upper mineral soil - not specifying soil depth. We now explain this in the caption of Tables 2 and 3.

COMMENT 1.25: *"Figure 1 - Specify what 'Species-averaged' means or reword so that the caption and figures are interpretable independent of the text. The range of SOC data (1-3%) is nowhere near the ranges presented in Figure 5 and 6 (0.1 to 100%), where the data is presented on a log-scale axis. Explain or resolve the discrepancy."*

This remark relates to the different SOC range for the Swedish database versus the range used in the IIASA-metric (the latter being much smaller). To clarify this issue, we now mention the different SOC axis under Figs. 7 and 8, and emphasize in these captions that SOC varied widely in the Swedish database.

We also rephrased the caption in Fig. 1 to make clear that the scores were averaged over all crop species.

GENERAL COMMENTS RAISED BY REFEREE 2

COMMENT 2.1: *"In the introduction the authors write that they aim at developing a globally valid metric while using data from Sweden only. It should be clarified that the upgraded metric is only valid for Sweden. Possibly, a global metric is not achievable at all, since nutrient availability is not limited by the same factors in different ecosystems worldwide. Hence, a metric for Sweden might be usable for other boreal, but not for tropical ecosystems. These considerations should be discussed."*

We agree with the referee that the metrics presented in the current manuscript are only valid for Sweden and have emphasized this in the revised manuscript (Lines 14, 91-93, 498-500, 530, 548-553). Future evaluations of the current metric against data from ecosystems (forests) outside Sweden are needed to verify its applicability, and to further update its formulation. We also agree that serious challenges are ahead, such as differential nutrient limitation and especially extrapolation to other ecosystem types. In the revised manuscript, we added a new Section 4.5 discussing the main challenges and provide an outlook for how to deal with issues such as differential nutrient limitation:

"Even though normalized productivity was significantly related to soil properties, and to our upgraded metrics, much of the variation in normalized productivity remains unexplained. The considerable unexplained variation may have multiple reasons. Apart from a possible lack of soil and nutrient data more closely related to N availability than the ones available in our database, another possible factor reducing R^2s could be the quality of the data in the database. This could for instance be due to an insufficient number of replicates sampled per data point (n = 3 for the soils), although this is probably of limited importance because of the large number of data points in the database itself. A more important source of uncertainty is probably the inevitable uncertainty related to the response variable, i.e. “climate-normalized” aboveground productivity. This not only includes uncertainty in the original productivity estimates (for which for example differences in management or disturbances likely increased variability), but 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 a consequence, the low R^2 values are party 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 randomly selected from our database for additional evaluation of our upgraded metrics.

Despite the limited R^2, similar significant results for different methods (1 and 2) and subsets of the database (regions, soil moisture classes and soil types) indicated that the findings about the soil properties and nutrients are generally robust. The upgraded metrics explained up to 21 % of the variation in normalized productivity. It is unclear to what degree the influence of nutrient availability is covered by this percentage. Future studies, where additional soil data can be included, will need to verify this. In any case, the significant relationships with normalized productivity, the better implementation of the soil variables and the capability of the metrics to explain up to 38 % of the variation in productivity across different gradients imply a significant improvement compared to the original IIASA-metric for this database.

A key challenge in the further development of a metric describing spatial variation in nutrient availability both within and outside the boreal biome is differential nutrient limitation. Eventually, we want to be able to compare for example N-limited and P-limited systems. The original structure of the IIASA-metric, which was kept in our upgraded metrics, facilitates this by allowing the inclusion of multiple soil variables such as C:N (mainly relating to N availability), pH (among others a critical factor controlling P availability) and exchangeable bases in one single metric. In fact, the IIASA-metric is particularly useful in this regard, as it gives more weight to the soil factor with the lowest score. This corresponds to reality and enables accounting for the type of nutrient limitation. For instance, if 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 2.2: *“The performance of the metric is bad. Instead of discussing the – non-existing – relationships, the authors should rather discuss the possible reasons for the failure of the metric. One possible reason is data quality. The authors should describe the soil sampling design and the methods used for chemical analyses. Inventory data might not be suitable to find relationships between parameters even though they exist, because (soil) variability would require a large number of replications, which is often not affordable in inventories.”*

We thank the referee for his/her critical assessment, which has helped us in interpreting the current metric’s performance and was very useful for developing future perspectives.

We realize that the predictive power of the upgraded metric was rather low. Data quality may to some extent contribute to the seemingly inadequate performance of the metric. However, we identified other reasons for the low R^2s that are probably more important than data quality. First, the necessary normalization procedure for climate increased the uncertainty on normalized productivity and lowered coefficients of variation. This became clear in our new analyses using randomly selected local nutrient availability gradients. Within such a gradient, normalization for climate was not necessary and the predictive power of the metric increased up to 38 %.

On the other hand, the real proportion of variation explained by nutrient availability is simply unknown and hence it is impossible to set a target for the R^2. Some of the unexplained variation is, for example, likely due to differences in management and disturbances. Given that our new analyses on the local gradients reveal that a substantial proportion of variation is explained by our upgraded nutrient metric, we consider these metrics a useful starting point for follow-up studies evaluating (and further upgrading) the metric.

In the revised manuscript, we added a new section (Section 4.5) about "sources of uncertainty and future challenges". We mention there data quality as a possible reason for the low R^2s, and explain that especially the normalization procedure also contributed substantially to the uncertainty (given the substantially higher R^2s for the local gradients where normalization was not needed).

As requested, we have also added more information about the forest plots, replicates etc. in the "Methods" section and in the caption and notes of Table 1. Finally, we refer to other papers that offer more detailed information about soil sampling and laboratory analyses in this table.

SPECIFIC COMMENTS RAISED BY REFEREE 2

COMMENT 2.3: *"17 - The coefficients of determination are that poor that you should not write "Normalized productivity increased with decreasing soil C:N ratio, while SOC exhibited an empirical optimum."*

We rephrased this sentence (Lines 16-17), and added a section to the discussion on sources of uncertainty, including reasons why R^2s were low (Section 4.5).

COMMENT 2.4: *"21 - The coefficient of determination of the upgraded metric is still poor and should not be called "a significant fraction".*

We rephrased this sentence to make it more careful (Lines 25-26).

COMMENT 2.5: *"34 and 36 - "among terrestrial ecosystems" and "global scale" is misleading. You should clarify that in the present paper only Sweden is considered."*

Even though we aim to further upgrade the metrics from this paper with data from ecosystems outside the boreal biome, we agree that in the previous version of our manuscript, it was not always sufficiently clear that only Swedish data were considered. In the revised manuscript, we therefore stress this at multiple places in the text (Lines 14, 91-93, 498-500, 530, 548-553), and rephrased the sentence indicated by the referee (Lines 43-44):

"Comparing nutrient availability among terrestrial ecosystems is thus difficult for two reasons: comprehensive and harmonized data on soil properties and nutrients are not usually available from experimental and observational sites, and no standardized quantitative metric exists to compare the nutrient statuses of terrestrial ecosystems at the global scale, or even at a national scale (e.g. for Sweden, which is considered in the present paper)."

COMMENT 2.6: *"47 - "are more indicative" to what and for what?"*

Soil properties such as soil texture, SOM and pH are more indicative of the general nutrient status, which is determined by the interplay between different soil nutrients and influenced by soil properties like pH. We rephrased this in the manuscript (Lines 54-57).

COMMENT 2.7: *"48 - What is meant with "the size of the soil solution"?*

We meant the "size of the soil solution pool [of a particular nutrient], exchange sites and unavailable soil pools (…)", and correspondingly clarified this in the manuscript (Lines 56-57).

COMMENT 2.8: *"57 to 60 Yes! Perhaps the fact that a global metric might not be achievable at all, since nutrient availability is not limited by the same factors in different ecosystems worldwide, should be discussed here, too."*

As answered to COMMENT 2.1, we agree with the referee that differential nutrient limitation poses a challenge in the development of a global nutrient metric. The retained structure of the original IIASA formula seems, however, particularly useful for tackling this issue. This metric allows including multiple soil properties and nutrients and gives more weight to the soil factor that is most limiting (i.e. the one with the lowest score). We incorporated our answer in the discussion of the revised manuscript (Lines 531-538):

"A key challenge in the further development of a metric describing spatial variation in nutrient availability both within and outside the boreal biome is differential nutrient limitation. Eventually, we want to be able to compare for example N-limited and P-limited systems. The original structure of the IIASA-metric, which was kept in our upgraded metrics, facilitates this by allowing the inclusion of multiple soil variables such as C:N (mainly relating to N availability), pH (among others a critical factor controlling P availability) and exchangeable bases in one single metric. In fact, the IIASA-metric is particularly useful in this regard, as it gives more weight to the soil factor with the lowest score. This corresponds to reality and enables accounting for the type of nutrient limitation. For instance, if 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 2.9: *"78 to 84 If the goal is a "global metric", data from the Swedish forest inventory service do not represent "a substantial variation in nutrient availability". Which were the additional variables? P?"*

We fully agree with the referee that the variation in nutrient availability across Sweden is much smaller than global variation in nutrient availability. The reason for using the Swedish forest inventor data was primarily the fact that this dataset met the main data needs, i.e., a database that combines the necessary information on soil properties and nutrients with data on plant productivity, while also covering a substantial variation in nutrient availability and containing a large number of sites to increase statistical power. We have clarified this in the revised manuscript (Lines 91-93) and also emphasize now in several places that the current metric applies to Sweden only and further analyses are needed to develop a globally applicable metric (Lines 14, 91-93, 498-500, 530, 548-553).

The additional soil variables referred to were soil total N concentration/stock and especially soil C:N ratio. We rephrased this sentence in the manuscript (Lines 89-91).

COMMENT 2.10: *"89 - Here you restrict your results on Swedish forests – you should already earlier mention that your goal is a metric for Sweden – not a global metric."*

Done (Lines 14, 43-44, 91-93).

COMMENT 2.11: *"94 to 134 - List all the parameters (soil, tree/productivity, climate/meteorology) and explain how they were measured."*

We listed the parameters in Table 1 of the revised manuscript, and mentioned the most relevant sampling, analysis, … details in the caption and underneath the table. We provided references for further details on soil sampling and analyses.

COMMENT 2.12: *"121 to 130 - Explain in more detail how the scores were derived and what you can find in the look-up tables."*

We rephrased the paragraph (Lines 130-137), and the caption of Fig. 1. We also added a hyperlink to the look-up tables.

COMMENT 2.13: *"137 to 138 - You should state that you call the two alternative ways to calculate normalized productivity "method 1" and "method 2" in the following."*

Done (Lines 148-159).

COMMENT 2.14: *"138 - Name the advantages and drawbacks."*

We now explain the main disadvantages of each method in this paragraph (Lines 155-159).

COMMENT 2.15: *"179 - Is "method 1" referring to the method used to calculated the normalized productivity?"*

Yes, it is. This should be clearer now that we rewrote the paragraph about the normalization methods (Lines 148-159, 168-176). We also referred to Fig. 3 here.

COMMENT 2.16: *"193 to 200 - You used half of the dataset from southern Sweden to upgrade the metric and tested the upgraded metric with the rest of the dataset including middle and northern Sweden, right? Did you also upgrade the metric using method 2 (half of the dataset for calibration, half for validation)"*

The referee's interpretation is correct. Thanks to this comment, we realized that this paragraph, where we explained the metric upgrades, was not sufficiently clear. We therefore rewrote the paragraph (Lines 218-232). We now also developed a second upgraded metric, based on method 2.

COMMENT 2.17: *"216 to 217 - Why not using both SOC and TN in separate models and choosing the model that better fits the data?"*

Using SOC always resulted in higher R^2s than using TN, irrespective of the normalization method on productivity (R^2 was 0.145 [method 1] or 0.048 [method 2] for SOC, while it was 0.116 [method 1] or 0.046 [method 2] for TN). We did not explicitly mention this in the manuscript, since we preferred to continue with SOC anyway because (i) it is a component of the IIASA-metric, and (ii) SOC/M better represents the organic matter content, which not only acts as a nutrient reservoir, but also provides cation and anion exchange sites. We clarified our *a priori* choice in the paragraph (Lines 261-264).

COMMENT 2.18: *"225 - "... related to normalized productivity (Table 2), however, the coefficient of determination was small (0,002 to 0,146)."*

Done (Line 293), and we also rewrote our assertions more carefully. See also the remarks on R^2s (COMMENTS 2.2, 2.3, 2.4, …).

COMMENT 2.19: *"227 - I cannot see from Fig. 5 that "the effect became more pronounced towards the south".*

With this sentence we mean that both the slope and R^2 increase from north to south (also shown in Table 2). We clarified this in the manuscript (Lines 294-295). See also COMMENT 2.31: we reassessed whether the R^2 in the north was lowest, and confirm that the information provided in the figure, table and text is correct.

COMMENT 2.20: *"232 to 234 - According to Table 3 also other variables than SOC, pH and C:N were included in the multiple regression models. Why?"*

Even though the relationships between the other soil variables and normalized productivity were less clear than for SOC, pH and C:N, they were also included by the model selection procedure. In other words, the minimum cross-validation mean squares were reached for models with additional soil variables, but the role of these variables in the simple and multiple regression formulas was less consistent.

COMMENT 2.21: *"239 to 240 - On the one hand you write that SOC and C:N "consistently describe a distinct, clear effect on normalized productivity" on the other hand you write that R2 is "at least a few percent" – this is contradictive. Both the figures and the R2 show that there is no distinct, clear effect on normalized productivity."*

We resolved this discrepancy in the paragraph (Lines 303-305):

"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 (Table 3)."

COMMENT 2.22: *"225 to 241 - This part should be rewritten since the variance in productivity is hardly explained by the soil variables (see also discussion l. 272 to 313)."*

We rewrote sentences in this part and elsewhere more carefully now (Lines 291-305). See also the notes on low R^2s in our replies (COMMENTS 2.2, 2.3, 2.4, 2.18) and in the manuscript (Section 4.5).

COMMENT 2.23: *"251 to 252 - The relationship between residuals and variables is very weak."*

If residuals of the (normalized) productivity-metric associations are still related to soil variables already included in the metric, this could point to a suboptimal implementation of the variables. We clarified this better in the manuscript (Lines 22-23, 161-163, 213-216, 312-318, 340-346, 349-351, 466-469). We also refer to our answers on the comments on low R^2s (COMMENTS 2.2, 2.3, 2.4, 2.18, 2.22 and Section 4.5).

COMMENT 2.24: *“254 - To my opinion, the results do not show “that SOC, C:N and pH are important factors influencing nutrient availability in Sweden.”*

We formulated this more carefully. The sentence now reads as follows:

“From the statistical analyses for *question 1*, we deduce that SOC, soil C:N and pH each play a role in influencing nutrient availability in Sweden.”

COMMENT 2.25: *“264 to 265 - From Figure 8 no enhancement of the upgraded metric can be deduced. The performance is still bad.”*

In contrast to the IIASA-metric (Fig. 9), both upgraded metrics are significantly positively related to normalized productivity across the database (Figs. 10 and 11). Moreover, our new analyses demonstrate that our metrics each explain (more) variation in productivity for five randomly selected nutrient availability gradients (Tables 4 and 5). The metrics are thus an improvement, at least for the Swedish boreal forests considered in the present study. They should not be seen as an endpoint, but rather as a first step in a development process.

COMMENT 2.26: *“272 to 313 - The variance in productivity is not explained well by the soil variables. Instead of discussing the (significant but very weak) relationships, you should rather discuss why you do not find relationships were you would expect them. What about the soil sampling design and the methods used for chemical analyses? You did not describe them in the material and methods section. Could the sampling design (number of replications) explain the bad model performance? How is the variability of soil variables in Sweden?”*

In response to the referee’s concern, we discuss in the revised manuscript the potential sources of uncertainty (<u>Section 4.5</u>). We also provide (references to) more information about sampling methods in the methods section (Table 1). See also our more elaborate reply to the COMMENT 2.2.

COMMENT 2.27: *“323 - ... but again with a very small coefficient of determination.”*

See other replies to comments on low R^2s (COMMENT 2.2, …).

COMMENT 2.28: *“330 - “... the nutrient availability metric was <u>intended to be</u> improved by ...”*

At least for the Swedish database, the upgraded metrics are an improvement compared to the IIASA-metric. See also other replies to comments on metric performance (COMMENT 2.25) and low R^2s (COMMENT 2.2, …).

COMMENT 2.29: *“335 - I cannot agree. The relationships were significant but only a very small part of the variance was explained by the variables used.”*

We rephrased the sentence to make it more careful. In the revised manuscript, we state that the upgraded metrics explained some variation (especially for gradients - Lines 482-493):

“In contrast to the original metric developed by IIASA, the upgraded metrics described some variation across all approaches using the full database (Figs. 10 and 11). Variables were generally properly implemented, at least for upgraded metric 1 (Table 5), while for metric 2, significant associations emerged between residuals of normalized productivity and SOC and pH (Table S17). These associations had, however, opposite trends depending on whether normalization method 1 or 2 was considered, meaning that the methods did not agree on whether the metric score was either under- or overestimated for the soil variable considered. The stratified analyses confirm that the metrics are an improvement, at least for those soil moisture classes and soil types with sufficient data points (Tables S12-15). Moreover, each metric could describe spatial variation in productivity for five randomly selected nutrient availability gradients (Tables 4 and 5). The coefficients of determination were generally higher for these gradients than for the database analyses, likely because the gradients did not require a normalization for climate (the latter increased the uncertainty on the response variable, see section 4.5 on sources of uncertainty and future challenges). Lastly, the gradients generally confirmed the correct implementation of soil variables in upgraded metric 1 (Table S18), whereas for metric 2, scores for high SOC might be overestimated (Table S19).”

See also the other replies to comments on low R^2s (COMMENT 2.2, …).

COMMENT 2.30: *“336 to 339 - Why are especially stable N isotope signatures and ion exchange resin bags of interest? What about other methods?”*

N isotope signatures and ion exchange resins are just two examples. We clarify this better in the revised manuscript and also mention a few other options (Lines 502-505).

COMMENT 2.31: *“Table 2 - Are you sure that R² is the same for all three regions for the parameters SOC, N stock, sand, clay and pHKCl? This is quite unrealistic and the data shown in Figure 5 (SOC) lead to the assumption that R2 of the North is largest and that of the South smallest. In line 226 to 227 you write that the effect becomes more pronounced towards the south (C:N), however, in Figure 5 the relationship is worse for the south than for the other regions and a larger R2 – as written in Table 2 – seems to be quite unrealistic.”*

In cases where R^2_{tot} was written instead of R^2, we presented outputs of ANCOVA models of which the cross-validation mean squared error was minimal when excluding a soil variable[(2)] x region interaction. Consequently, results for all three regions were based on one model. When R^2 was written, there was a (significant/selected) soil variable x region interaction, and splitting up the analyses among regions yielded separate parameter estimates (here, we could have provided a joined R^2_{tot} as well, while slope and intercept estimates would have remained the same, but splitting up provides more information on exactly these R^2s).

We reassessed whether the R^2 in the north was lowest, and confirm that the information provided in the figure, table and text is correct.

```
Call:
lm(formula = Resid[Region == "N"] ~ CNup[Region == "N"])

Residuals:
     Min       1Q   Median       3Q      Max 
-2.03337 -0.51938  0.01981  0.53435  2.52545 

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)    
(Intercept)          0.343074   0.104818   3.273 0.001131 ** 
CNup[Region == "N"] -0.013609   0.003858  -3.527 0.000455 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7218 on 545 degrees of freedom
  (64 observations deleted due to missingness)
Multiple R-squared:  0.02232,	Adjusted R-squared:  0.02052 
F-statistic: 12.44 on 1 and 545 DF,  p-value: 0.0004555
```

```
Call:
lm(formula = Resid[Region == "M"] ~ CNup[Region == "M"])

Residuals:
    Min      1Q  Median      3Q     Max 
-3.2659 -0.7302  0.0153  0.7555  3.9562 

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)    
(Intercept)          0.692602   0.151206   4.581 5.40e-06 ***
CNup[Region == "M"] -0.026695   0.005439  -4.908 1.12e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.117 on 780 degrees of freedom
  (67 observations deleted due to missingness)
Multiple R-squared:  0.02996,	Adjusted R-squared:  0.02871 
F-statistic: 24.09 on 1 and 780 DF,  p-value: 1.12e-06
```

```
Call:
lm(formula = Resid[Region == "S"] ~ CNup[Region == "S"])

Residuals:
    Min      1Q  Median      3Q     Max 
-5.3262 -0.7532  0.1444  0.9772  6.3101 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)          2.04169    0.18359   11.12   <2e-16 ***
CNup[Region == "S"] -0.08163    0.00742  -11.00   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.441 on 949 degrees of freedom
  (110 observations deleted due to missingness)
Multiple R-squared:  0.1131,	Adjusted R-squared:  0.1122 
F-statistic: 121.1 on 1 and 949 DF,  p-value: < 2.2e-16
```

COMMENT 2.32: *"Table 4 and 5 - The coefficients of determination are similar and do not point on a better implementation of the parameters in the upgraded metric."*

This comment presumably concerns a misunderstanding. Here, we examined the implementation of the soil variables in the metric. If residuals of the (normalized) productivity-metric associations are still related to soil variables already included in the metric, this could point at a suboptimal implementation of the variables. We clarified this better in the manuscript (Lines 22-23, 161-163, 213-216, 312-318, 340-346, 349-351, 466-469) and in the captions of Tables S9 and S16-19. We also refer to our answers on the comments on low R^2s (COMMENTS 2.2, 2.3, 2.4, 2.18, 2.22 and Section 4.5).

For upgraded metric 1, only few relationships between residuals of normalized productivity and the soil variables were significant (Table S16), while the opposite was true for the IIASA-metric (Table S9). As now discussed in the manuscript (Lines 483-487), residuals of normalized productivity were significantly related to both SOC and pH for upgraded metric 2, but the sign of the relationship depended on the normalization method (Fig. 3), i.e. methods 1 and 2 did not agree on whether partial scores for SOC and pH would be over- or underestimated. Analyses of the gradients generally confirmed the correct implementation of soil variables in upgraded metric 1 (Table S18), whereas for metric 2, scores for high SOC might be overestimated (Table S19).

COMMENT 2.33: *"Figure 1 - I have problems understanding the legend of Figure 1."*

We rephrased the caption of Fig. 1.

TECHNICAL CORRECTIONS SUGGESTED BY REFEREE 2

COMMENT 2.34: *"12 - …to test which combination of soil factors…"*

Done. (Line 12)

COMMENT 2.35: *"63 to 64 - rephrase this sentence to avoid the twofold use of "recent(ly)"*

Done. (Line 69-71)

COMMENT 2.36: *"96 - explain the abbreviations"*

Done. (Line 104-105)

COMMENT 2.37: *"118 - mass stock [kg m-2]; if really the mass is meant, the formula is wrong"*

Corrected. (Line 127)

COMMENT 2.38: *"157 - "we therefore we split" delete one "we"*

Done. (Line 173)

COMMENT 2.39: *"164 - "SOC" – stock or content? "Total N" – total N content? "N stock" – total N stock?"*

We clarified this in the manuscript. (Line 179-180)

COMMENT 2.40: *"167 - Name the software used."*

Done. (Line 181)

COMMENTS RAISED BY THE ASSOCIATE EDITOR

COMMENT 3.1: *"Reviewer 1 makes the excellent suggestion to stratify before you do your analyses. This reviewer specifically suggests to stratify according to wetness class before exploring soil fertility x productivity relations. I think that this is an important suggestion and deserves a more careful consideration than simply doing the same analysis for four different site wetness classes. I would even argue that you can only expect soil fertility to directly influence productivity if the site is well drained. Of course the IIASA approach does not consider this since agricultural sites are preferable located in better drained sites (unless it is rice), but with forests this is obviously different. Since you have a large database, you should consider to explore the suggestion to stratify further and explore whether you find stronger relations if you stratify according to soil type. At least you are then trying to find relations for sites with comparable soil forming factors."*

We followed the advice of referee 1 to stratify the main analyses by soil moisture, and we have also stratified by soil type (separately), as suggested here by the editor. In the revised manuscript, we added a section 'Identifying potentially confounding factors', in which we first present how soil properties and normalized productivity are related to soil moisture and soil type (Sections 2.3.1, 3.1 and 4.1). In the next sections we then present the stratified regression analyses (on soil variables + N deposition) and evaluations of the upgraded metrics. The regression outputs (Tables S3-S6 and S12-S15) generally confirmed the robustness of our results. We briefly discuss outcomes of the stratified analyses in the manuscript (e.g. Lines 306-308, 337-340, 487-488, 524-526).

COMMENT 3.2: *"Reviewer 1 also suggests to include N deposition in your analysis. Your write 'Nonetheless, if the editor so wishes, we are willing to test the influence of N deposition separately and present it in the revised manuscript`. My answer is: yes, I wish that you analyse whether N deposition affects productivity. It is something that you should have done after this suggestion was made and before writing your answer. Please do not forget to stratify here as well, since I predict that site wetness and other soil forming factors will interact with any N-deposition x productivity relation. There is a wealth of literature showing how strong the effect of N deposition is on forest productivity and just ignoring this by declaring it is out of the scope of this study is not helpful if N deposition affects productivity in your dataset."*

Both referee 1 and the editor emphasized that the role of N deposition should be discussed. We therefore investigated the influence of N deposition on soil variables, productivity, and on climate-normalized productivity (e.g. Lines 189-198, 280-289, 391-402). As explained on Lines 280-289 in the manuscript, N deposition indeed showed a strong positive relationship with productivity, because productivity increases in the north-south direction, as do N deposition, temperature and incident light. Whether N deposition correlated with normalized productivity depended on the normalization procedure (Fig. 3): for method 1 (which normalized productivity for species and climate [+ indirectly N deposition!]), there were mostly no positive associations with N deposition, while for method 2 (actual/attainable productivity of spruce), there was a strong positive relationship. This positive relationship followed logically from the increase of the ratio actual/attainable productivity from north to south Sweden. These issues, and their consequences, are discussed in the revised manuscript on Lines 391-402:

"Many studies have shown the strong influence of N deposition on forest productivity (e.g. Binkley and Hogberg, 2016; From et al., 2016). As expected, N deposition correlated to some extent with some of the soil variables considered in the present study, such as the total soil N stock and concentration (Fig. 4b). Furthermore, N deposition was strongly positively related to productivity. However, this effect of N deposition on productivity cannot be separated from the influence of climate and light, as all these factors increase together in the north-south direction. Nevertheless, we argue that for the goals of this study, i.e. investigating soil nutrient-productivity relationships across Sweden and developing a nutrient metric, the spatially varying N deposition is not problematic, since the normalization for climate and species according to method 1 (Fig. 3a) at the same time also removed the influence of the confounding N deposition on productivity. Accordingly, "Residual productivity" was generally not correlated with N deposition (Table S3). The response variable derived from method 2 (i.e. actual/attainable productivities for spruce - Fig. 3b) in contrast, correlated strongly with N deposition (Table S4), because both actual/attainable productivity and N deposition increased from north to south. Consequently, relationships between actual/attainable productivity and soil data for this method were unavoidably confounded by N deposition."

COMMENT 3.3: *"While I understand your arguments why you chose the IIASA approach, it would still be interesting if a purely statistical approach would lead to better results. Why don't you present this as well (and don't forget to stratify here as well)?"*

In our revised manuscript, we present results from multiple regression analyses (Table 3) and compare them with our upgraded metrics (Eqs. (11–13), (14–16) and (10)) in the discussion (Lines 494-505, also see COMMENT 1.3):

"Variation in normalized productivity explained by the upgraded 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 can, however, easily be updated, while equations derived from multiple regressions would be of less use for building a nutrient availability metric because such an entirely empirical approach does not allow for later updates of parameters or model structures based on data from other ecosystems. In order to further enhance performance of the metrics, and to test to what extent they can describe variation in nutrient availability outside of Swedish conifer forests, additional datasets with productivity and soil information are needed. (…)"

COMMENT 3.4: *"Reviewer 2 points at the limitations of your study for developing a global metric. I agree with this reviewer that this is not possible but I do think that a metric for e.g. 'boreal forests on well drained soils' would already be a major improvement compared to the present situation. I think you should address this more specifically in your revised manuscript and acknowledge that a global metric is not realistic."*

Both referees and the editor expressed doubts on whether developing a global metric of nutrient availability is a realistic goal. We fully agree that this is an ambitious objective that might not be fully attainable (e.g. due to too large differences among vegetation types). However, no one has ever made a serious attempt. We thus consider it too early to "acknowledge that a global metric is not realistic", and have therefore kept the two main aims of our paper, which were to (i) detect important soil variables related to nutrient availability in Sweden, and (ii) start the development of a nutrient availability metric.

In any case, even if development of a global metric turns out too complicated, a valuable alternative could be to develop a metric for (boreal) forests only for example. In response to this comment and related COMMENTS 1.5, 2.1 and 2.8, we added a new Section 4.5, discussing the main challenges, and provide an outlook for how to deal with issues such as differential nutrient limitation, e.g.:

"A key challenge in the further development of a metric describing spatial variation in nutrient availability both within and outside the boreal biome is differential nutrient limitation. Eventually, we want to be able to compare for example N-limited and P-limited systems. The original structure of the IIASA-metric, which was kept in our upgraded metrics, facilitates this by allowing the inclusion of multiple soil variables such as C:N (mainly relating to N availability), pH (among others a critical factor controlling P availability) and exchangeable bases in one single metric. In fact, the IIASA-metric is particularly useful in this regard, as it gives more weight to the soil factor with the lowest score. This corresponds to reality and enables accounting for the type of nutrient limitation. For instance, if 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 3.5: *"Reviewer 2 also makes the excellent suggestion that data quality (including comparability of data potentially analyzed with different chemical methods) may be an issue. In your answer you are not addressing this but it is still a very valid consideration. Were uniform methods used? How many laboratories were involved in the data? Is anything reported about detection limits? How are missing data treated? These are all valid questions that you have to address in your manuscript. You cannot expect people in dig into the original dataset to find this out."*

Based on this suggestion and the related COMMENT 2.2, we added a new Section 4.5 about "sources of uncertainty and future challenges" in the revised manuscript. We mention data quality there as a possible reason for the low R^2s, and explain that especially the normalization procedure also contributed substantially to the uncertainty (given the substantially higher R^2s for the local gradients where normalization was not needed).

As requested, we have also added more information about the forest plots, replicates etc. in the "Methods" section and in the caption and notes of Table 1. Finally, we refer to other papers that offer more detailed information about soil sampling and laboratory analyses in this table.

COMMENT 3.6: *"In you answer you refer twice to the need for using datasets where climate does not vary but nutrient availability does (e.g. local gradients). I was wondering why are you not exploring this? If you have a dataset with 2500 points I am sure that there are selections that can be considered local gradients. In my opinion you have the data to test this; it is not something that should be postponed to the future."*

We thank the editor for this excellent idea, and have now evaluated our upgraded metrics against data from five nutrient availability/productivity gradients in Sweden. These gradients were selected to cover a range in nutrient availability and productivity within the same/very similar climate. The results from these analyses indicated that tree productivity was significantly related to our metrics for all gradients (Tables 4 and 5). Moreover, the R^2s of these relationships were generally higher than was the case when considering the complete database (R^2s were up to 38 %). This suggests that the (necessary) procedure to normalize productivity for factors like climate in the previous analyses reduced the predictive power of the metrics. This last aspect also relates to referee 2's comment on the low R^2's, which we have now addressed in a section on sources of uncertainty and future challenges in the discussion (Section 4.5).