

Subject: Reply to the main Comments of Ref#2

Here, we present replies to the main comments of Referee #2. Point-by-point responses to all General, Specific and Technical Comments and changes in the manuscript will be presented in a later stage of the peer review process.

Sincerely,

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

## **REPLIES TO GENERAL COMMENTS**

We thank the reviewer for the insightful comments. We agree that the nutrient availability metric in its current form cannot just be applied to ecosystems globally, and that the development of a global metric is challenging, e.g. in part because of differential nutrient limitation across ecosystems. We also agree that the low  $R^2$  of the upgraded metric indicates that it may not be good enough yet to describe a desirable proportion of the variation in nutrient availability, even for boreal forests, but this remains to be tested with other datasets. Instead, it should be seen as a starting point in a development process. We will discuss the type of data needed for further development and evaluations of the metric in the manuscript. For the soil properties and nutrients, however, we still find that, despite the limited  $R^2$ , our findings are generally robust (see below). In the discussion, we will therefore add a section on robustness of the associations we found (also see the reply to Ref#1's comments).

COMMENT: *“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 that a metric developed or improved based on data from Swedish forests alone cannot be extrapolated to be used as a global nutrient metric. We will make this more explicit in our manuscript. Even though our ultimate aim is indeed to develop a globally applicable metric, the metric upgrade presented in this manuscript should only be seen as a first step in the development process, i.e. a step where we indicate the need for such a metric and show the development approach with its advantages and disadvantages. Furthermore, we will add a section in the discussion on additional data needed to develop a globally applicable metric, such as data from local gradients in nutrient availability (which have the advantage that no normalization for climate is needed, see below). Evidently, further updates of the metric based on data from other ecosystems are also needed, not only because the proportion of the variation explained is rather low, even within Sweden (see below), but also because, as Referee #2 mentions, different ecosystems can be limited by different nutrients, while the current metric was developed based on only N limited forests.

Although we agree with Referee #2 that differential nutrient limitation across ecosystems poses an additional challenge to the development of a nutrient metric, we anticipate that the implementation of both N and P availability and perhaps other nutrients into one single metric is not necessarily an impossible task. In theory, it is perfectly possible to include multiple 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 and could therefore account 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. Based on an example like this one, we will discuss the challenge of differential nutrient limitation in the manuscript.

COMMENT: *“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.”*

First, the low  $R^2$ s lead us to ask the question how large they should be in order to sufficiently describe variation in nutrient availability. We agree with the referee that the performance of the upgraded metric is likely still too low for applications (although it can already describe > 20% of the variation in southern Sweden), but we should be cautious with using  $R^2$  as the main criterion to evaluate the metric: the unexplained variation depends on multiple factors not directly related to nutrient availability that could not be accounted for in our analyses (e.g. imperfect productivity estimates and the normalization procedure add to this unexplained variation, see below). Instead, we also considered the significance of the relationships, and the presence or absence of a remaining association between residuals and soil variables in the metric as additional criteria for testing its performance. Besides these criteria, the metric should obviously also succeed in explaining variation in complementary datasets. We thank the reviewer for this critical assessment, which inspired us to include in the revised manuscript a paragraph discussing the criteria of a good nutrient availability metric.

Besides factors such as management that increase variation, we think that the low  $R^2$  values also follow from (i) a lack of soil and nutrient data more closely related to N availability (the primary limiting nutrient in these boreal forests), and (ii) inevitable uncertainty related to the response variable, i.e. “climate-normalized” aboveground productivity. The first point implies that variables such as C:N may be insufficient. This requires further testing with datasets that contain multiple indicators of N availability (not available for our dataset). We will briefly discuss the potential of a few other N availability indicators in the revised manuscript. The second point includes uncertainty in the original estimates of productivity, and uncertainty related to its normalization for climate: by taking residuals of the productivity vs climate regression model, we for instance unintentionally not only remove the direct effect of climate on productivity, but also its indirect effect on productivity through nutrient availability. In contrast, the approach taking actual/attainable productivity as a response variable does not suffer from this issue, but the estimates of attainable productivity are uncertain. As a consequence, low  $R^2$  values are not just a result of variables failing to explain variation in the real nutrient availability, but also because the normalization procedure has shortcomings that can only be overcome by using datasets where climate does not vary but nutrient availability does (e.g. local gradients).

Finally, although the relationships between normalized productivity and soil variables and nutrients have low  $R^2$  values, the significant relationships still point at the role of these variables

in shaping nutrient availability. Especially the observations that for example SOC and C:N appear as significant factors, across both normalized productivity approaches and across soil moisture classes (see the reply to Ref#1's comments), confirm the robustness of our statements. Hence, we do not agree that the relationships are "non-existent", but we do agree that certain claims on for instance the importance of the variables should be described with more caution, and that the use of the upgraded metric remains to be evaluated, even within the boundaries of Sweden.

We agree with Referee #2's main remarks on generalizability of the current metric and on the fact that the difficulties in its development should be discussed. We will thus process the answers given in this reply in the manuscript, and make further clarifications and corrections based on the Specific and Technical Comments.

---