Response to referee #1

Thank you very much for your elaborate review of our manuscript. We appreciate the time and effort put into this extensive review and your in-depth comments. We also very much appreciate the recognition of our work as a potential important contribution to the literature on forest growth responses to drought. Below, we will address your questions and comments individually in four sections.

The first question is about the time span of the ERA5 climate data and the remote sensing data (e.g. MODIS EVI) being different and how we dealt with this difference in the models. The answer is that we did not use the remote sensing data in the models as only the ERA5 climate data was used as a product with a temporal dimension (see also next section in this response). Besides the different time spans being an obstacle for using the remote sensing data in the model, keeping the remote sensing data out of the models also ensured that we could more confidently compare the remote sensing data and model predictions without lapsing into circular reasoning. In the revised version of the manuscript we will make sure that this will be explained more explicitly in the methods section.

Concerning the second issue raised, about using variables without a temporal dimension, such as soil properties, plant traits (e.g. SLA) and biomass, we acknowledge that these properties are not constant. We also agree that this might become a problem when these properties have dramatically changed in the past decades such as due to deforestation. This is a limitation of our analysis that deserves to be discussed more elaborately in the discussion section. However, as the referee acknowledges, the products used often provide only one value (e.g. SLA) without a time dimension and others are indeed not available when going back to the 1980's and 1990's, such as the high resolution biomass maps. Furthermore, when checking the site locations in our database prior to the analysis, we found that most if not all of the sites are (still as of 2020) located in protected areas and nature reserves. Visual inspection of these sites using recent aerial images (e.g. Google Earth) also shows that these inventory sites, while sometimes located in a very small forest fragment, seem to be relatively undisturbed. Nonetheless, for the revised version of the manuscript we have looked into the data sets recommended by the referee (Hansen et al. 2013 or Song et al. 2018) and included the long-term vegetation continuous field (VCF) dataset (1982-2019) from Song et al. 2018 into our analysis. In both the new leaf litterfall as well as the new stem growth models, the VCF was chosen as one of the 20 variables used in the final models. Therefore, it is likely explaining some of the temporal and spatial variability in stem growth and leaf litterfall data that was not explained by the ERA5 climate data and other geospatial datasets. We appreciate your suggestion of including the VCF dataset in our analysis.

We agree that our current validation approach does not validate model performance across the sites and that we do not currently know how well the models perform over space alone, only over space and time combined. Therefore, we welcome the idea from the referee to include a second validation of model performance by splitting the test and training data based on the sites and see how well the model performs at the sites not included in the training data. To test how well the model performs across the sites, we have trained the model using the predefined selected features and the hyperparameters (from the existing model) but now only using training data consisting of measurements from 60% of the sites in the dataset. The model evaluation on the data from the remaining 40% of the sites is presented (Figure 1) and will be included in the revised supplement. As expected, the model performance declined in both the stem growth as well as the leaf litterfall models when using only 60% of the sites, more so in the leaf litterfall model than the stem growth model. The absolute (RMSE) and relative error (NRMSE) increased only by about 0.01 to 0.04 Mg C ha^{-1} month^{-1} and 1 to 4.4 percentage points, respectively. However, the explained variability (R^2) did decline significantly from 0.5 to 0.4 in the stem growth model and from 0.67 to 0.38 in the leaf litterfall model. These results suggest that the model performs better when explaining mainly the temporal variation in leaf litterfall and stem growth compared to mainly the spatial variation. The results also suggest that the model performs better in predicting the spatial variation in stem growth compared to leaf litterfall, which might indicate that some of the drivers of spatial variability in leaf litterfall are not well represented in our current model.

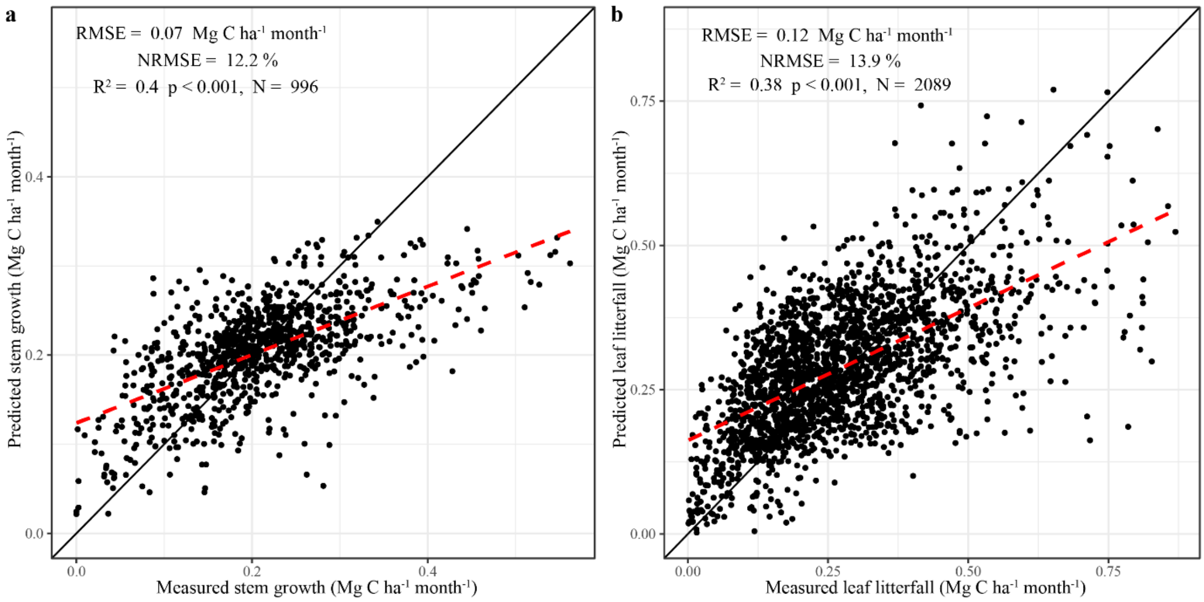

Figure 1 Model evaluation of the new site based validation. The scatterplots show the predicted biomass production versus the measured biomass production of the test data (40% of the field sites) that was used to validate the stem growth (a) and leaf litterfall (b) models. The dashed black line is the 1:1 line and the dashed red line the least squares linear regression fit.

The XGBoost model unfortunately does not offer a way to easily obtain uncertainties from the model output. The structure of the model, which is a sequence of regression trees and not a group of separate regressions trees that are averaged like in a random forest model, does not allow for a standard uncertainty estimate. There are to our knowledge a few workaround techniques to still obtain uncertainty estimates from XGBoost models. One of these workarounds is to train a separate XGBoost model to predict the error in the original test dataset (see https://medium.com/@qucit/a-simple-technique-to-estimate-prediction-intervals-for-any-regression-model-2dd73f630bcb). We employ this method in the revised version of our manuscript to derive a measure of uncertainty for the stem growth and leaf litterfall estimates (Figure 2). The results of this new uncertainty analysis are depicted in Figure 2 and will be included in the new supplement accompanying our revised manuscript. The figure shows that the absolute error (RMSE) of the stem growth and leaf litterfall models is low in the high elevation ecosystems of the Andes compared to the lowlands (Figure 2a, 2b). However, when the RMSE is normalized using the average seasonal range in stem growth and leaf litterfall values (annual amplitude) the opposite pattern is observed (Figure 2c, 2d). In this case, the relative error for both the stem growth and leaf litterfall models is very high in the

Andes region (> 40%) compared to the lowland forest of the Amazon basin, the Cerrado, Caatinga, and Atlantic forest regions and central America (< 20%). These results suggest that the performance of both models is relatively weak in mountainous ecosystems, presumably due to large differences in climate and soils on relatively short spatial scales, compared to the other ecosystems included in the analysis. Thank you for your suggestion to perform an uncertainty analysis as it will increase the usability of our model output.

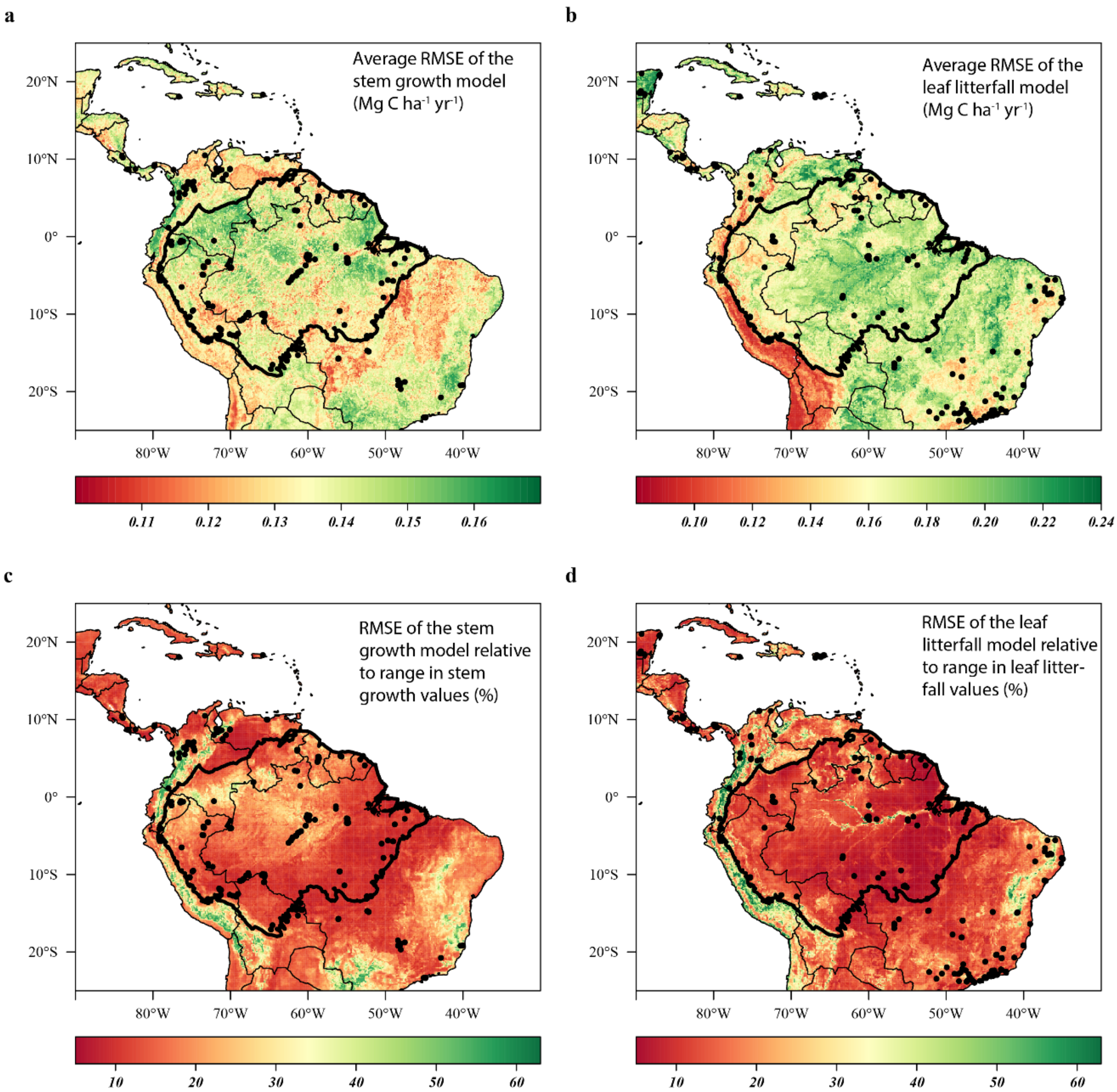

Figure 2 Absolute and relative errors of the stem growth and leaf litterfall XGBoost models. The monthly RMSE of both models was averaged from 1982 to 2019 (a, b). The relative error was calculated as the averaged RMSE divided by the average seasonal range in stem growth and leaf litterfall estimates (c, d).

Thank you again for your elaborate review and we hope that with this response we have addressed your most pressing concerns and questions regarding our manuscript. We are confident that addressing your major and minor comments in the revised version of the manuscript, the quality of the manuscript will be greatly enhanced.

On behalf of my co-authors,

Thomas Janssen