Response to referees

We were happy to read these very positive, thoughtful reviews of our Discussion Paper. We have prepared a revised MS taking the three referees' comments into account. The comments are reproduced here, in full, with **bold type** used to highlight the key points requiring consideration. Our responses to every comment are given in *italics*.

Anonymous referee #1

General comments

This paper addresses the issue of trait variability within plant functional types in relation to climate. The authors argue that most traits vary in relation to climatic gradients but that change occurs sometimes through trait change within PFTs, and sometimes through shifts in the occurrence and abundance of PFTs with innately different trait values. These ideas are not new and have been suggested and shown in various earlier papers, but the authors present an impressive dataset that is in my opinion more reliable than those upon which earlier analyses have been based. Other tests of trait-environment relationships at similar spatial scales often rely on data collated from multiple studies and environmental data of poorer quality than that used in this manuscript.

The manuscript is very well written and a pleasure to read. The research questions are clearly stated and clearly addressed by the data and analyses. I have few major criticisms, and most of my comments are merely suggestions that I believe would improve the quality and readability of the manuscript and do not point to major flaws in the methodology or interpretation.

There is **only one major issue** that I would like to see addressed in the manuscript before publication. My problem with most studies investigating trait-environment relationships is that despite identifying statistically significant relationships among environmental predictors and trait means, the ability to predict trait values from environmental forcing alone is often quite poor. That is, after accounting for the variance explained by the environmental predictors, and differences among PFTs, most of the trait variability still remains (there is overlap in the traits values observed at either extreme of the environmental gradient and in almost all PFTs). Given that vegetation models are attempting to simulate vegetation and biogeochemistry with these same environmental forcings alone – which account for so little of the observed trait variation – I think it is important to acknowledge limits to the predictability of the vegetation traits we actually observe when using only macroclimate and a few simple soil variables to force models and discuss other potential sources of variation (e.g. microclimatic variation, disturbance, heterogeneity, evolutionary constraint). Specifically I would like to see a **paragraph highlighting these issues** and some attempt at **variance partitioning** in the analyses that highlights the proportion of trait variation explained by climate, PFTs, and unexplained variation.

These points are very well taken. We have added a new paragraph at the end of the Discussion that covers this topic (plus a new sentence at the end of the Abstract). We have also taken the opportunity to present a variance-partitioning analysis, which we think is a significant enhancement to the manuscript.

I would also really like to see another plot, similar to figure 2, but with **PFTs plotted in trait space** to highlight to massive overlap in trait values among PFTs.

We have included a new Figure (Fig. 5), which illustrates exactly this point.

Specific comments

Table 1 seems unnecessary. With most of the important information contained within it repeated in figure 1. I would suggest moving it to the Supplementary info.

We have moved the Table as suggested.

DGVMs typically use A_{max} and V_{cmax} as parameters to describe photosynthetic rates. I understand that chlorophyll florescence offer advantages over A_{max}/V_{cmax}, but perhaps the authors should state this because otherwise it seems that it would have been more appropriate to measure parameters used by DGVMs.

V_{cmax} can be inferred from A_{max} using the so-called one-point method, but even simply measuring A_{max} requires taking a heavy infrared gas analysis (IRGA) apparatus to the field sites. Fluorescence is a different measure, not a substitute for A_{max}. However, we sacrificed the added value of providing A_{max} and V_{cmax} data for the ability to assemble a very large trait data set, including sampling forest reserves in South China that would have been challenging to reach with an IRGA.

It is not stated in the methods what the sampling unit is that is being analyzed. Are the GLMs applied to trait values of individual plants and the environmental variable at the site levels or (as is plotted in figure 3 and 4) species-level means?

The sampling unit analysed is a species at a site. We have now added this information in the Methods.

I would be interested to know how the analyses look when they are repeated within and between species rather than PFTs. In terms of trait change along environmental gradients this is more relevant as species are real entities rather than PFTs which are the invention of modelers. There isn't room for this in this manuscript, which is focused on the utility of traits and PFTs for models, but I look forward to seeing this analysis in anther publication.

We agree that there is scope for this analysis to be done in future, and that it would be of great interest. However we agree that this manuscript doesn't have room for it; it would be a separate study.

Does the entire PCA analysis of climate space in China and the climate space covered by the sites only exist to make the point that the sites cover a wide range of climate conditions that are representative of Chinese climate? This point seems to be adequately made in figure 1. I would suggest moving this analysis to the supporting information.

Yes, this is the only purpose of this complete PCA. So we agree, and we have moved this analysis out of the main text (apart from a brief mention of its similarity to the analysis of our sampling sites).

Please add some discussion of the confounding role of nutrients. Many of the traits included in this analysis would be expected to be profoundly altered by different soil nutrient conditions. It is understandable given the poor quality of most soil nutrient maps that this was not included as a predictor variable in the analysis, but I believe further discussion and an acknowledgment of their role should be included.

We fully agree; we have added new text in the Discussion, with appropriate references, making this clear.

Technical corrections

In the discussion it is stated: "On the other hand, the LPJ-family models treat SLA as a PFT-specific parameter and thus do not allow for covariation of SLA with N_{area}." This is not true. Please see the recently published LPJml-FIT model:

Sakschewski, B., von Bloh, W., Boit, A., Rammig, A., Kattge, J., Poorter, L., Peñuelas, J. and Thonicke, K. (2015), Leaf and stem economics spectra drive diversity of functional plant traits in a dynamic global vegetation model. Global Change Biology. doi: 10.1111/gcb.12870

We have modified this statement, and cited this paper.

Anonymous referee #2

General comments

This manuscript addresses a scientifically interesting and important topic – the connection between biophysical, biochemical and physiological leaf traits and climate. It is asking the question how temperature and water availability influence selected leaf traits, and if observed variability in leaf traits is due to continuous adaptive trait shifts within plant functional types (PFTs), or due to PFT replacements along climatic gradients. Although many contemporary Dynamic (Global) Vegetation Models (DGVMs) still rely on the PFT concept, more recent developments in vegetation modeling focus on PFT-less trait-based approaches, arguing that the conventional PFT-based approaches may be too rigid, general, and over-simplifying. As PFTs are usually assigned a fixed set of parameter values in many DGVMs, continuous adaptive variation cannot be modeled directly, but needs to be mimicked by PFT replacement, disregarding that within-PFT variability of traits can equal or exceed between-PFT variability. Especially with respect to the more recent development of trait-based DGVMs that allow continuous trait variation within defined ranges of trait space this study provides valuable new data material for model parameterization, calibration and evaluation. Moreover, it contributes to an improved understanding of the linkages between different leaf traits and climate.

The abstract is clearly structured, presenting the motivation for the presented study, the methods used to address the problematic, a very short summary of the main results found in the study, and a brief outlook on the significance of the presented results. The introduction provides background information on the PFT concept in vegetation modeling and its inherent drawbacks, and establishes a connection to related studies in the field. However, **although the scientific questions acting as motivation for this study are inherently present hidden in the introduction text, I would like to**

see them listed and phrased directly as such (bullet points, listed by numbers, or in a similar way) at the end of the second paragraph (p. 7097, l. 21).

We have made this change: now the third paragraph starts by stating the main scientific questions addressed, as a numbered list.

The scientific methods used to address the research questions are well-established and appropriate to address the presented research questions. Results and conclusions are generally presented in a transparent, structured and concise way and sufficiently illustrated with figures and tables. The discussion part is well-structured, but section 4.3 (Comparison with previous studies of trait variation within and between PFTs) is rather short and basically only focuses on Kattge et al. (2011) and Zhang et al. (2012). If available, it would be desirable to (shortly) include a few more studies related to this topic here.

The literature on this topic is actually quite limited; we have found a couple of additional papers to cite, but both use the examples of N and SLA and rely on a graphical demonstration – they add little to the analysis by Kattge et al. However, we realize that the original section looked unbalanced, so we have re-structured it, added the new references, and altered the section heading to a more general one.

The only part where I do not fully agree with the authors is the last section (4.4 Implications for modelling). Their results nicely show that within-PFT trait variability is often continuous and can be as high or higher as between-PFT trait variability, and that PFT replacement only partially explains trait variation along climate gradients. Nonetheless, they strongly argue in favor of PFT-based vegetation models, when in fact their results justify and support the development and existence of PFT-less trait-based models that allow dynamic plant community assembly in dependence of prevailing environmental conditions. I would have liked to have seen a slightly **more balanced discussion here with respect to the advantages/limitations of PFT-based vs. trait-based vegetation models**, as in my opinion both model types have their right to exist, depending on the research questions under consideration.

This was a thought-provoking comment. On reflection, we realized that although simply retaining PFT distinctions would be one way to account for differences between PFTs, in fact the phenomenon described (trait-environment relationships being contingent on PFT membership) could as well or better be represented as an effect of 'PFT-defining traits' on the relationships between other traits and environment. We have amended our discussion accordingly.

In general, the manuscript is sufficiently referenced to allow making connections to related research. In a few places where more references would be desirable, it is highlighted in the "Detailed Comments" section of this review. The provided supplementary material is of good quality and sufficient to transparently present the results in a more in-depth way than possible in the manuscript itself. Overall, I consider this study to be a valuable contribution to the field, and suggest its publication in BG after minor revisions.

Detailed comments

p. 7096, l. 19-22: " Usually a fixed set of properties (parameter values) is assigned to each PFT. This expedient simplifies modelling, but it is a potential weakness because it disregards continuous adaptive variation within PFTs and the possibility that such variation is "universal" – that is, manifested similarly within and between species, PFTs and communities." Should phrase this even more clearly: the definition of a limited number of PFTs with fixed parameter values is an artificial generalization concept used by vegetation modelers to discretize continuous trait combinations into a manageable number of seemingly distinct categories. However, this oversimplification neglects that the range of trait variations within these artificial PFT-categories in reality may be as large or larger than between PFT categories, which leads to an underestimation of the plasticity and adaptive potential of vegetation to environmental change and vegetation feedbacks to climate.

We fully agree. We have strengthened our statement on the subject.

p. 7097, l. 10-21: this part reads a bit confusing, as the arguing in favor of trait-based approaches vs. PFT-based approaches goes back and forth and from the phrasing is not clear enough. Please first talk about the advantages of trait-based modeling, and then make it clear that afterwards you are talking about circumstances in which leaf traits have been discovered to be distinctly different between PFTs, thus indicating that PFT-based modeling also can be justified based on focus and circumstances. "An advantage of trait-based modelling is that it can take better advantage of the wealth of georeferenced data now available on plant functional traits (Kattge et al., 2011)." This may indeed be one advantage, but in my opinion the more relevant advantage of trait-based modeling is that it allows to simulate continuous trait variation, thereby allowing the development of plant community assemblies that are adapted to site-specific biotic and abiotic environmental conditions and can react more flexibly to environmental change, as well as allowing new approaches to simulate functional diversity and competition (see, e.g., the trait-based aDGVM2 model as described in Scheiter et al., 2013).

We agree, and we have added a statement to this effect.

"On the other hand, some leaf traits can have different relationships to climate depending on the PFT": I'm not entirely sure I understand correctly what you mean to say with this sentence. Do you mean that certain leaf traits within one PFT behave one way along a climate gradient, while they behave the opposite way along the same climate gradient for plants that belong to another PFT, or do not vary with climate at all for a third PFT? If possible, rephrase this to make it more clear.

We have clarified the wording of this statement.

p. 7097, l. 25: "on all of the species present at 80 sites, with a wide geographic spread.": How many species did you sample overall? And what was the range of species numbers between sites (minimum and maximum number of species per site)?

This information was available in Table S2. However, we have now extracted key statistics (number of species-site combinations; range of species number) and put them into the text.

p. 7098, l. 2-4: "Area-based nutrient contents provide no independent information, as they are simply derived from mass-based nutrient contents and SLA, but they provide an alternative perspective on the regulation of leaf nutrient contents." In what way do they provide an alternative perspective? Please elaborate a bit more closely, or else it becomes hard to justify why you are presenting both if they are not independent.

We have added a senetence expanding on this point, and provided a key reference in support of our reasoning.

p. 7098, l. 12: "The sites (Table 1) represent...": Please also point out Figure 1 here, as you show the location of your sampling sites on a map in Figure 1. I find it helpful to also see the location of the sites on a map, and would have asked for a map figure, but then realized that the sites are actually highlighted in Fig. 1 when looking at the figure later on.

Done.

p. 7099, l. 11/12: "except for a few species with very small leaves at the driest sites": What did you do for these? No sampling, or sample as much as was available?

We took a smaller sample (at least 2 g). This is now stated.

p. 7099, l. 19: "Leaf C was measured by the potassium dichromate volumetry method and leaf N by the microkjeldahl method." Please add a reference if possible, as it is not instantaneously clear to everybody how these methods work.

Done.

p. 7103, l. 7/8: "Figure 1 also shows the frequency of different GDD0-α combinations among grid cells, and the site positions in this climate space.", and p. 7121, Fig. 1: frequency distribution and location of sampling sites in climate space: The chosen sampling sites cover a large range of the occurring GDD0-α combinations and are therefore well-suited for the pursued study purpose. However, no sites cover the very low end of GDD0 values, and in the area of GDD0-values between 6000 and 9000 and α between 0.4 and 0.9. Which areas of China would these combinations correspond to? Probably high-altitude grid cells for the low GDD0-values, and places in Southern China with high GDD0 and intermediate to high α values? Maybe point this out briefly either in the Figure caption or in the text.

We have provided this information in the revised text.

p. 7103, l. 14-16: "Deciduous trees and deciduous shrubs favor cooler and drier climates, corresponding to the deciduous forests of central eastern China." I generally agree with your four PFT groups based on optimum and tolerance thresholds, except maybe for this group, as their tolerance range compared to the ones of the PFTs in the other groups is very wide, and therefore makes these two PFTs overlap with group (1) and group (3).

Their range is particularly wide, certainly, and we have now commented on this.

p. 7104, l. 3-6: "Inclusion of PFTs as predictors (Fig. S1 in the Supplement) shows that there are some differences among PFTs in the typical trait values found at any

given α. This is most obvious for biophysical traits – LA, SLA and LDMC – and area-based nutrients." Are the differences statistically significant? (Same question applies to Fig. S3)

Yes. The original text mentioned the fact that we applied a quite conservative (P < 0.01) criterion in order to minimize the chance of "false positives" in analyses (2) and (3), where a large number of parameters are estimated. However, it was stated kust once and without elaboration. So we have now expanded this statement in the Methods, explained that by "significant" we always mean P < 0.01 (or better), and provided frequent "reminders" of this in the Results section.

p. 7104, l. 25/26: "Warmer climates also show somewhat reduced potential and actual quantum yield." So both dry conditions and warm climate show reduced QY. Can you make a judgment which of these two factors has the greater effect? I suppose that, since dry conditions and warm conditions are not statistically independent, the decrease in QY is a combination of both, but nonetheless it would be interesting to know more about the relative importance of each factor.

As should now be clear (from our more explicit treatment of "significance", QY shows independently significant effects of both drought and heat. We can't say whether one is more important than the other: that would depend on the conditions sampled.

p. 7105, l. 24/25: "The observed continuous biophysical trait variations with moisture availability are consistent with previous studies..." Please add some references for these studies.

This was an oversight. We have now added a list of relevant papers.

p. 7108, l. 24-28: "Our findings suggest that vegetation models should retain the PFT concept and a minimal set of PFTs, because the distinctions between woody and herbaceous, de- ciduous and evergreen, and angiosperm and gymnosperm plant types systematically influence the values of key biophysical traits in ways that would not be predictable from assumed universal relationships." I do not agree with this statement. Vegetation models using the PFT concept may be useful and sufficient to address many scientific questions with respect to vegetation dynamics, but it is not true that using the PFT concept is the only way to get clear distinctions between key biophysical traits. Trait-based vegetation models not necessarily need to assume universal relationships, but may define a potentially allowed maximum range of values for key biophysical traits. If plants are assigned values from within these ranges at birth, selection through environmental conditions such as temperature and water availability will lead to the emergence of successful trait value combinations, whereas unsuccessful combinations will be eliminated through competition and/or environmental pressure. Whether woody or herbaceous, deciduous or evergreen, plants with low or high SLA, etc. will prevail therefore is not predefined, but will be an emergent property of simulated trait selection through environmental filtering. No PFTs need to be predefined, but an *a posteriori* classification of simulated plants into PFT categories based on simulated successful trait combinations is possible and will also pick up PFT replacement over time and space where it occurs. Trait-based vegetation models such as, e.g., the aDGVM2 model (Scheiter et al., 2013) therefore offer completely new approaches to simulate changes in functional diversity, trait

selection through environmental conditions, and competition for resources (water, light) that in such a way are not possible with classical PFT-based vegetation models.

See our response to the related general comment previously. Although a broad discussion of model construction principles is outside our scope here, we accept the point and have modified our statements accordingly.

p. 7122, Fig. 2: Change "boardleaf" to "broadleaf in figure caption.

Done.

What are the grey background points? The GDD0 vs. α combinations of all the 10 km grid pixels in China?

Presumably this comment refers to Fig. 2? We have amended the caption to point out that the grey points are, indeed, all of the grid cells across China.

p. 7123, Fig. 3: Change "boardleaf" to "broadleaf" in figure caption

Done.

p. 7124, Fig. 4: Change "boardleaf" to "broadleaf" in figure caption

Done.

Supplementary Material: Please also change "boardleaf" to "broadleaf" in figure captions where applicable.

Done.

Anonymous referee #3

This article provides an impressive dataset and very insightful analysis of the modulation of plant functional traits within and between functional groups according to environmental conditions broadly related to moisture availability and temperature. Although several studies on this subject exist, few can provide such a large, uniformly measured dataset together with high quality environmental data. The analysis is aimed at disentangling the effects of adaptive trait shifts within functional types (PFT) versus PFT replacement along environmental gradients on trait variability. This question is extremely relevant not only to improve our understanding of functional trait modulation in general but also due to the (still) frequent and arguably problematic use of the PFT concept in dynamic vegetation models.

Specific comments

The paper is very well written and generally easy to follow. Apart from some minor points, which are outlined in "technical corrections", I have only three **main points of critique**.

Firstly, the relative lack of discussion of the issues related to the use of plant functional groups, such as inferring functional similarity in groups where trait syndromes vary widely and **functional group membership may be dependent on**

environmental conditions (e,g. species generally known as growing in tree form growing in shrub-like forms in harsher environments). Although this is addressed to some extent, e.g. on page 7096, I think such a fundamental issue should be highlighted even more. This is particularly important since the conclusions seem to point towards the authors interpreting their study as being supportive of the PFT concept, when their results could easily be interpreted otherwise.

This is a good point. The revised Discussion draws attention to this further limitation of the PFT concept.

Related to this, **I miss references to species-specific, or at least genus specific trait differences**. Again, this is briefly mentioned in the discussion but without providing any data or analyses. Although repeating the analyses performed on species rather than functional group level might change the scope of this work too much, at least giving an indication to the degree of within as opposed to between species functional variability in their data would add a lot to the paper.

As this referee notes, consideration of trait differences between taxa, as opposed to PFTs, would require different analyses and change the scope of the paper in a major way. Our goal in this MS was to compare variation within and between PFTs (basically a modelling construct), rather than addressing the separate question of trait plasticity within species. But this question is relevant to modelling too; so we have added a little more Discussion material on it.

Secondly, the argumentation for using GDD0 instead of e.g. the Principle Component scores of axes one or MAT, which has a higher loading than GDD0 on the first axis, is insufficient. There are good reasons for using GDD0 but the authors should clarify these.

We have added a sentence explaining why we use GDD_0. The main argument in favour of GDD_0 as opposed to MAT is that in climates with cold winters, such as in northern China, large spatial variations in MAT can be induced by conditions in midwinter that do not relate to conditions during the time when growth occurs.

Also, since a lot of traits are highly dependent on soil nutrient status and other environmental conditions, which have not been measured in this study, their likely effect should at least be mentioned.

This is now mentioned in the Discussion, as a potential field for investigation which we have not attempted.

Finally, unless I have overlooked this, no statistical tests of differences between the linear model regression fits have been performed. The authors talk about "significant differences" (e.g.pp. 7104 line 16) but it is not clear how significance can be inferred without such tests.

See our response to the same point as made by referee #2. Indeed, all statements made in the original manuscript had been carefully checked for significance using a conservative (P < 0.01) criterion. We hope that this is now abundantly clear in the revised MS.

Technical corrections

Pp. 7099 - line 11: clarify what you did in case of the very small leaves at the driest sites.

Done (also requested by referee #2).

Pp. 7099 - line 15: how many leaves were scanned?

We have included this information now.

Pp. 7100 - line 14: the results of the country-wide PCA should be provided in the appendix.

Done (also recommended by referee #1).

Pp. 7101 - line 1: clarify how you divided the variables into bins - how did you decide on the size of the bins?

This is now briefly described.

Pp. 7101 - line 6: GLM should be GZLM to avoid confusion between General linear models (GLM) and the Generalized linear models (GZLM) used here.

We disagree. GLM is a very widely accepted abbreviation for generalized linear models; whereas we have never encountered "GZLM". Given the context, we believe there should be no confusion.

Pp. 7105 - line 25: please provide references to the "previous studies".

Done (also requested by referee #2).

Figure 1: if possible, please choose another colour scheme to cater for red-green blindness, it would be very useful if you could code the plot symbols according to region

Done.

Please check your spelling of "broadleaf" in the figure captions.

Corrected.

It would be useful to include a table with the number of species per site to give the reader an idea of differences in the geographic spread of the species you measured and whether the PFTs within certain regions are represented by many or few species.

This information is already in Table S1. We have also provided summary statistics in the revised text.

The authors