We would like to thank reviewer 2 for the positive comments and good suggestions. We respond in detail to the reviewer's comments below (our responses are given immediately below the reviewer's comments).

Major comments:

1. Why didn't the authors use the information on soil age at the chronosequence sites for a spin up which is consistent with the age of the soils? Such a test would be a much better test for the calibration of the model than the presented test using equilibrium simulations.

Response: We agree that a more strenuous test would be starting the model from parent material and simulating the long-term (4.1 million years) soil and ecosystem development along the chronosequence. However, it is not feasible to do that in this study because of the immense amount of computational time required. In addition, realistic driver data is not available for the long geological time period.

The following statement has been added to address this concern (Page 14, lines 401-406).

"One limitation for the simulations along the Hawaii chronosequence is that we prescribed the soil P pools using site measurements rather than running the model from parent material to simulate soil and ecosystem development. A simulation over the time period of 4 millions of years is infeasible for CLM as it requires unrealistic computation time. Realistic driver data is also not available for this long time period."

2. The model is heavily calibrated and the parameter values differ strongly from site to site. From the current analysis it is unclear which processes and thus which parameters are responsible for the occurrence of phosphorus limitation. A parameter perturbation experiment would be very beneficial to gain a better understanding of the models dynamics.

Response: We acknowledge that the model is heavily calibrated because of limited observational data on P dynamics and C-N-P interactions. We used a single set of parameters for the five Amazon sites so the simulated variations in NPP across the five sites are due entirely to differences in the climate condition and soil properties, not to differences in parameters. A different set of parameters was used for the P-limited Hawaii chronosequence site because we used site measurement of C: P ratios in leaf, leaf litter, fine root, wood, and soil for model parameterization.

Following Zaehle's suggestions in his comments, we have run the simulation at the P-limited site using the leaf C:P ratio of the N-limited site and the simulated NPP is only slightly higher (about 3%). This suggests that P limitation at this site is mainly caused by

soil P availability rather than the leaf C:P ratio. In addition, the experiments we have shown in Fig. 7 also suggest that ecosystem responses to elevated CO2 strongly depend on the parameters that control soil P availability.

We agree that a sensitivity analysis would help improve the understanding of the model. However, a comprehensive sensitivity analysis for CLM requires significant amount of computation time and is beyond the scope of this study. We will pursue this in our future study. The following statement has been added in the revised manuscript to address this (Pages 19-20, lines 575-577).

"Additionally, a comprehensive sensitivity analysis would help us to improve our understanding of the model and quantify model uncertainties, which will be an important part of our future research. "

3. As the author correctly state the P and N cycles are close linked, thus it would be very interesting to the see (1) the results from the N limited Hawaii sites in Fig. 2 and (2) the simulated NPP for the CN simulation in Fig.7a.

Response: The results for the N limited Hawaii site is now added as Fig 2a. In Fig. 7, the purpose of the three experiments is to show how the two processes controlling P availability in CLM-CNP (both of which could be enhanced in response to increasing CO2, as shown in Fig. 6) might affect NPP, soil P pools and carbon storage under elevated CO2. Therefore we feel it is not appropriate to add CN simulation results in Fig. 7.

4. It is not clear to me why the model performance greatly improved by the introduction of the P cycle based on the results shown in Fig4b. Statistical measures would be beneficial in this case.

Response: We use Mean Absolute Percent Error (MAPE) to measure the model performance. MAPE is 7% for CNP model and 15% for CN model. This has been added in the text. The following text has been added (Page 14, lines 411-412):

"MAPE (Mean Absolute Percent Error)(Kothamasu et al., 2004) is 7% for CLM-CNP and 15% for CLM-CN."

5. The author state that the P cycle can be substituted with the N cycle (P14455, L8). This view seems problematic to me, as the mechanisms underlying N and P limitation differ strongly. A substitution might not work when the temporal evolution of the NPP response to an increase in CO2 is simulated.

Response: We agree with the reviewer. We have deleted that statement and added the following statement (Pages 15-16, lines 449-453):

"The present-day similarity between the extent of N limitation in CLM-CN and P limitation in CLM-CNP, however, does not imply that N limitation is a good surrogate for P limitation in the future. As discussed previously, N cycle and P cycle have different characteristics and their response to changes in atmospheric CO2 and climate will also be quite different."

6. The model description and the rationale behind the calibration strategy needs clarification:
- Missing is the description of P uptake and what assumption are used. -
The scaling parameter of the biochemical mineralization affects the CP ratio of the soil. How is it parametrized? The factor controls, together with the parameters for "Soil C:P", the actual simulated CP ratio of SOM. There is data on soil CP ratio. Can you use it to evaluate the parametrization? - The C:P ratios of SOM is much higher at the Amazon site compared to the Hawaii sites. What is the reasoning for changing the the C:P ratios rather than k_BC? - The parametrization of the leaching flux is not described. Together with the rate constant for conversion of secondary to occluded P, these rate constant control the amount of P in steady state. A description and perhaps a sensitivity analysis (see above) would improve the understanding of the model.

Response: Plant P uptake is determined by downregulating plant P demand from soils with f_{plant} (see section 2.1.3). f_{plant} , the overall nutrient limitation factor, is calculated in section 2.1.2. Plant P demand ($F^P_{plant\_demand}$) is the amount of P needed for the allocation of new growth to various tissue types based on specified C:P ratios for each tissue type and allometric parameters that relate allocation between various tissues (see Appendix A). Plant P demand from soil is the difference between total plant P demand and the P flux from the retranslocated P pool (section 2.1.1).

The scaling parameter of the biochemical mineralization is from Goll et al. (2012) and we have added the reference in the text.

The C:P ratios of SOM at the Hawaii sites are from the site observations. The C:P ratios at the Amazon sites are based on model calibration. With the C:N ratio of 10 to 12 in CLM, N:P ratios are between 41-50 in CLM-CNP, which is consistent with observations. Yang and Post (2011) reported that the mean N:P ratio in highly weathered tropical soils is 40.34.

The leaching loss of P depends on the concentration of soluble P in soil water solution and the rate of hydrologic discharge from the soil column to streamflow (CLM Tech note). Since we have solution P explicitly represented in the model, there is no parameterization involved for P leaching.

We agree that a sensitivity analysis would help improve the understanding of the model. However, A comprehensive sensitivity analysis for CLM requires significant amount of

computation time and is beyond the scope of this study. We will pursue this in our future study. The following statement has been added in the revised manuscript regarding this comment (Pages 19-20, lines 575-577).

"Additionally, a comprehensive sensitivity analysis would help us to improve our understanding of the model and quantify model uncertainties, which will be an important part of our future research. "

Minor comments:
P 14440, L6 : It's not a given fact the P is the most limiting nutrient, please rephrase.

Response: we have rephrased the statement as "currently believed to be the most limiting nutrient"(Page 1, line 12)

P 14440, L 26: It is not clear to me which results presented in this study justify this conclusion.

Response: The simulations with doubling CO2 suggest that tropical forest responses to increasing $[CO_2]$ will interact strongly with changes in the P cycle. Therefore, the predictive modeling of these interactions will be important for better prediction of future carbon uptake and storage in tropical ecosystems.

P 14441-14442: Most of the cited studies are missing in the "references" section, like Zaehle, Clark, Foley, Melillo, ...

Response: The references have been added.

P 14441, L 16: Change "Zahle" to "Zaehle"

Response: It has been revised.

P 14441, L18: It's not a given fact the P is the most limiting nutrient, please rephrase.

Response: It is now revised as "has been considered as" (Page 2, line 56).

P 144443, L 6: There is evidence that it is a Mo effect rather than a N effect (Baron et al., 2009). It should be mentioned that the N effect rather uncertain and other theories exist.

Response: We agree that there is evidence suggesting the importance of phosphorus (P) and/or molybdenum (Mo) availability in regulating N fixation. The following text has been added (Page 3, lines 72-74):

"There is evidence suggesting the importance of phosphorus (P) and/or molybdenum (Mo) availability in regulating N fixation ((Hungate et al., 2004;Reed et al., 2013;Barron et al., 2008)."

P14443, L 9: The parametrization of the Langmuir equation in CASA-CNP and in JSBACH are based on measurements using the Hedley fractionation (see Wang et al., 2010). Please explain the difference to your strategy. In particular, why using a soluble, labile and sorbed pools compared to a labile and a sorbed pool is an advantage.

Response: We first used the Uncertainty Quantification (UQ) package (Sargsyan et al., 2013) to narrow down the range of parameters needed in the model. Next, based on the measurements at the calibration site (NPP, vegetation C and P pools, and soil P pools measured using Hedley fractionation method), parameters in the Langmuir equation and in other equations controlling the transformations of soil inorganic P were optimized through a more traditional "trial and error" iterative calibration procedure. This has been described in section 2.4.2.

We explicitly consider soil solution P pool in CLM-CNP for the following reasons: (1) we assume plants can only take up P in soil solution; (2) in CASA-CNP and JSBACH, the assumption for describing the relationship between labile P and sorbed P using Langmuir equation is that these two pools are in equilibrium within days. For the half-hour time step in CLM-CNP, Langmuir equation will not be applicable to describe the relationship between theses two pools. We assume that solution P is in equilibrium with labile P pool for the time step we use (30 minutes) and therefore the relationship between solution P and labile P can be described using the Langmuir equation in CLM-CNP; (3) explicitly modeling of solution P is needed to model P leaching flux. We have added the following statement regarding this point (Page 6, lines 171-176).

"Additionally, solution P dynamics is explicitly modeled in CLM-CNP for the following reasons: (1) we assume plants can only take up P in soil solution; (2) for the time step we use (30 minutes), the relationship between solution P and labile P can be described using the Langmuir equation; (3) explicitly modeling of solution P is needed to model P leaching flux.

P14443, L 12: A higher temporal resolution can be an improvement but it is not per se. Please explain

Response: We have deleted this statement as Dr. Zaehle suggested that both other models can operate on half-hour time step.

P14444, L 11: The sentence is misleading as the diagram summarize P in plant

compartments in a single pool. It doesn't show the compartments „leaves", "fine roots",etc.

Response: The statement has been revised (Page5, lines 141-146).

 "we consider phosphorus in plant biomass (leaves, fine roots, live wood, dead wood, live coarse root, dead coarse root, and fine root), plant storage pool, coarse woody debris, three litter pools, four soil organic matter compartments, and five soil inorganic phosphorus pools including solution P, labile P, secondary mineral P, parent material P, and occluded P.

P14444, L13: In this study representation of mineral P different from Wang et al, 2010 is used: three pools (soluble, labile and secondary) which interchange P with each other compared to two pools in Wang et al., 2010 (labile and sorbed). Please explain your rationale to have these dynamics. I am a bit hesitate to accept that a more complex representation of inorganic P is an improvement when the process understanding is poor and data to parametrize such a model is scarce.

Response: The representation of soil inorganic P pools in CLM-CNP is consistent with previous studies on measurements using the Hedley fractionation method (Hedley and Stewart, 1982;Tiessen and Moir, 1993;Yang and Post, 2011), allowing for direct model parameterization and evaluation using measurement data. The reasons for explicitly modeling solution P pool have been discussed above. In the revised manuscript, we have discussed in detail our rationale to have a representation of soil inorganic P pools that is different from previous studies (Page 6, lines 160-176).

 "The representation of soil inorganic P pools in CLM-CNP is different from that of existing CNP modeling approaches (Wang et al., 2007; Wang et al., 2009; Goll et al., 2012). The five inorganic P pools we implement in the model are based on the measurements using the Hedley fractionation method (Hedley and Stewart, 1982;Tiessen and Moir, 1993;Yang and Post, 2011). The Heldey fractionation method has been well established as a comprehensive way to measure soil P (Smeck, 1985;Johnson et al., 2003;Cross and Schlesinger, 1995). The newly expanded global Hedley P database (Yang and Post, 2011) and the increasing number of soil P measurements using Hedley fractionation method will be helpful for model parameterization and evaluation. Our recently published global maps of various forms of P (using the same terminology) will provide a reasonable initialization for future applications of CLM-CNP on the global scale (Yang et al., 2013). Additionally, solution P dynamics is explicitly modeled in CLM-CNP for the following reasons: (1) we assume plants can only take up P in soil solution; (2) for the time step we use (30 minutes), the relationship between solution P and labile P can be described using the Langmuir equation; (3) explicitly modeling of solution P is needed to model P leaching flux."

P 14452, L 7: I guess, the parameter was tuned in both simulations, CN and CNP. (If not, the comparison of the C stocks would be rather unfair.) Please clarify this in the text.

Response: The parameter was not tuned for the CN model. The comparison is between the CLM-CNP and the original CLM-CN model so we did not tune the CN model at all. We have added the following statement for clarification (Page 13, lines 378-379):

"The higher SOM C in CLM-CNP is due to the adjustment of the specific decomposition rate of the slowest SOM pool."

P 14466, Table 3: "Soil C:P" is four times in the table

Response: they should be SOM1 C:P to SOM4 C:P. The table has been revised.

P 14472, Figure 6: The presentation of the feedbacks is incomplete. There is an arrow missing from "Phosphatase activity" to "available P". Phosphatase activity was shown to be depend on P availability in the field. It is not clear to me why "P demand" and "P supply" are grouped together. When the grouping is removed all signs of feedbacks could be given: There is a positive feedback from "P supply" to NPP based on the assumption P is the most limiting nutrient in tropical forests. P demand has a negative feedback to "available P" based on the assumption vegetation satisfies its demand by uptake.

Response: The reason that we group "P demand" and "P supply" together is to show that P limitation on NPP depends on the balance between P demand and supply. We have removed the oval shape that groups P demand and supply together in order to avoid any confusion for the readers. We have added the arrow between "Phosphatase activity " to "available P" in Fig. 6.

P 14456, L 25: It is not clear how the simulation were performed. Was the desorption rate / biochemical mineralization rate enhanced during the spinup or were the rates enhanced beginning with the CO2 increase?

Response: The enhanced rate was applied beginning with the CO2 increase because under elevated CO2, these rates could be enhanced due to the feedbacks shown in Fig. 6. We have revised the text to make it clear (Page 17, lines 501-507).

"In specific, we first ran the model using historical atmospheric [CO_2] until 2009, and starting from 2010 we increased atmospheric CO2 concentration to 800ppm(roughly doubled compared with 390 ppm in 2009) and continued to run three model simulations for 40 more years: control (desorption rate and specific biochemical mineralization rate kept unchanged), enhanced biochemical mineralization (biochemical

mineralization rate doubled), enhanced desorption (desorption rate doubled)."

Reference:

Barron, A. R., Wurzburger, N., Bellenger, J. P., Wright, S. J., Kraepiel, A. M., and Hedin, L. O.: Molybdenum limitation of asymbiotic nitrogen fixation in tropical forest soils, Nature Geoscience, 2, 42-45, 2008.

Cross, A., and Schlesinger, W.: A literature review and evaluation of the. Hedley fractionation: Applications to the biogeochemical cycle of soil phosphorus in natural ecosystems, Geoderma, 64, 197-214, 1995.

Hedley, M., and Stewart, J.: Method to measure microbial phosphate in soils, Soil Biology and Biochemistry, 14, 377-385, 1982.

Hungate, B. A., Stiling, P. D., Dijkstra, P., Johnson, D. W., Ketterer, M. E., Hymus, G. J., Hinkle, C. R., and Drake, B. G.: CO2 elicits long-term decline in nitrogen fixation, Science, 304, 1291-1291, 2004.

Johnson, A. H., Frizano, J., and Vann, D. R.: Biogeochemical implications of labile phosphorus in forest soils determined by the Hedley fractionation procedure, Oecologia, 135, 487-499, 2003.

Kothamasu, R., Shi, J., Huang, S. H., and Leep, H.: Comparison of selected model evaluation criteria for maintenance applications, Structural Health Monitoring, 3, 213-224, 2004.

Reed, S. C., Cleveland, C. C., and Townsend, A. R.: Relationships among phosphorus, molybdenum and free-living nitrogen fixation in tropical rain forests: results from observational and experimental analyses, Biogeochemistry, 1-13, 2013.

Smeck, N.: Phosphorus dynamics in soils and landscapes, Geoderma, 36, 185-199, 1985.

Tiessen, H., and Moir, J. O.: Characterization of available P by sequential extraction, in: Soil sampling and methods of analysis, 2 ed., edited by: Carter, M., and GREGORICH, E., CRC Press, Taylor & Francis Group, Boca Raton, FL, 75-86, 1993.

Yang, X., and Post, W.: Phosphorus transformations as a function of pedogenesis: a synthesis of soil phosphorus data using Hedley fractionation method, Biogeosciences, 8, 2907-2916, 2011.

Yang, X., Post, W., Thornton, P., and Jain, A.: The distribution of soil phosphorus for global biogeochemical modeling, Biogeosciences, 10, 2525-2537, 2013.