Dear Authors,

3 reviewers have assessed your revised manuscript. As you will see, not all reviewers were satisfied with the revision.

Two main points that stand our from the review reports: 1. the novelty of the work, 2. the evaluation of the presented model.

About the novelty of the work we can have a long debate (also the reviewers have different opinions here), but in my opinion in principle the addition of the P cycle in a major land surface mode (like JULES), is worth to be published as a model description paper in a journal like GMD. However, I advise the authors to be very clear and honest about novel and non-novel aspects in the text. If the P cycle is implemented in the model based on existing concepts, or other models this should be explicitly stated in the text.

The real weakness of the current manuscript for me is still the evaluation of the model. If this manuscript aims to be the reference study presenting the CNP version of JULES, then a more thorough model validation needs to be added to the analysis. I can agree that this would be a site-scale evaluation (i.e. not including a large scale regional evaluation). But it should at least be a multi-site evaluation, and preferably include an evaluation for an experiment (one of the reviewers is suggestion to use the Gigante nutrient addition experiments for example). Only with such an evaluation the manuscript can act as a reference publication for other studies that use the validated JULES-CNP model for applications and to address actual research questions.

A revised version of the manuscript should thus account for these two point and should address all other remarks and suggestions raised by the reviewers.

best regards,
Hans Verbeeck

Dear Hans Verbeeck,

We thank you for your comment on our submitted manuscript. Following are our responses and the modifications we did based on your two main points:

1- *Novelty of model:*

For the development of JULES-CNP which will be eventually incorporated in the UK earth system model, instead of coming up with brand new equations or processes that have not been yet incorporated in any global model which will need a lot of testing, we opted for implementing existing and already tested equations from global land surface/vegetation P enabled models. The current version of JULES-CNP forms the basis of future developments. As requested by reviewer 3 after the first round of reviews, we included citations to all equations taken from the literature.

This is now clearly stated in two parts in the manuscript, in the introduction in lines 119-123:
"Here, we describe the development and implementation of the terrestrial P cycle in the Joint UK Land Environment Simulator (JULES) (Clark *et al.*, 2011), the land component of the UK Earth System Model (UKESM), following the structure of the prior N cycle development (Wiltshire *et al.*, 2021) and utilising state of the art already tested and implemented descriptions of P cycling in other land surface models (Wang, Houlton and Field, 2007; Zhu *et al.*, 2016; Goll *et al.*, 2017)."

And also, at the start of section 2.2 on JULES-CNP description in lines 168-170:
"JULES-CNP includes the representation of the P cycle in JULES version (vn5.5) and it is built on existing and well tested representations of P cycling in other global land surface models (Wang, Houlton and Field, 2007; Yang *et al.*, 2014; Goll *et al.*, 2017; Sun *et al.*, 2021)."

However, a unique feature of our extended P component in JULES is the estimation of the soil organic and inorganic P sorption based on the saturation status of the relative adsorbed P pools. This is now clarified in the manuscript in the introduction (line 125 -129) as follows:

"The model (JULES-CNP) is parameterized and calibrated using novel in situ P soil and plant data from a well-studied forest site in Central Amazon near to Manaus, Brazil with soil P content representative of 60% of soils across the Amazon basin. The new developed P component estimates the sorption of the soil organic and inorganic P based on the saturation status of the adsorbed P pools, which is unique compared to the other existing P models and enable more realistic estimation of P ad/desorption processes."

2- *Evaluation:*

We have performed additional site-level evaluation to show model performance at other sites. The extended test sites are located in the Amazon (AGP-01, SA03 and CAX) which include a gradient of fertility from west to east Amazon, and two manipulation experiments one in the Gigante Peninsula in Panama and one at the Hawaii chronosequence (Hawaii Kokee). These site level simulations which were parameterised with site level tissue and soil C:P ratios and maximum sorbed P capacities using site specific parameters, showed a significant improvement of JULES-CNP over the C and CN only versions. Specifically, simulated C pools and fluxes with JULES-CNP were closest to the measurements as opposed to JULES C and CN which overestimated all observations at all test sites. Additional text in all sections (Abstract, introduction, methods and results) is included in track changes in the manuscript. Below we include tables with site selected (Table 3 and 5 in the text) and figures with results obtained (figures 4 and S8 in the text)

Table 3. Test sites name, location and climate characterises.

Site	**Name**	**Location**		**Climate**	
		Lat.	Lon.	Rainfall (mm yr^{-1})	Temperature(°C)
Study site	AFEX project	-2.58	-60.11	2431	26
AGP-01	Agua pudre plot E	-3.72	-70.3	2723	25.5
CAX	Caxiuanã flux tower site	-1.72	-51.5	2314	26.9
SA3	Tapajós flux tower site	-2.5	-55	1968	26.1
Gig. Pen.	Gigante peninsula (control data)	-9.1	-79.84	2600	26
Hawaii K.	Hawaii Kokee (control data)	22.13	-159.62	2500	16

Table 5. Additional test sites data used for model parameterisation

	AGP-01[a,b]	**CAX**[a,b]	**SA3**[a,b]	**Gig. Pen.**[c]	**Hawaii K.**[b,d]
$Leaf_{C:P}$	600	600	600	700	691.5
$Root_{C:P}$	1000	1000	1000	1750	1100
$Wood_{C:P}$	3000	3000	3000	5500	5937.5
$Soil_{C:P}$	2000	2000	2000	800	2000
K_{or-max}	0.001	0.001	0.001	0.0033	0.001
K_{in-max}	0.001	0.001	0.001	0.0185	0.001

[a]C:P ratios from Wang, Law and Pak, 2010 and [b]maximum sorbed P capacities from Yang *et al.*, 2014.
[c]Mirabello *et al.*, 2013 [d] C:P ratios from Vitousek, 2004

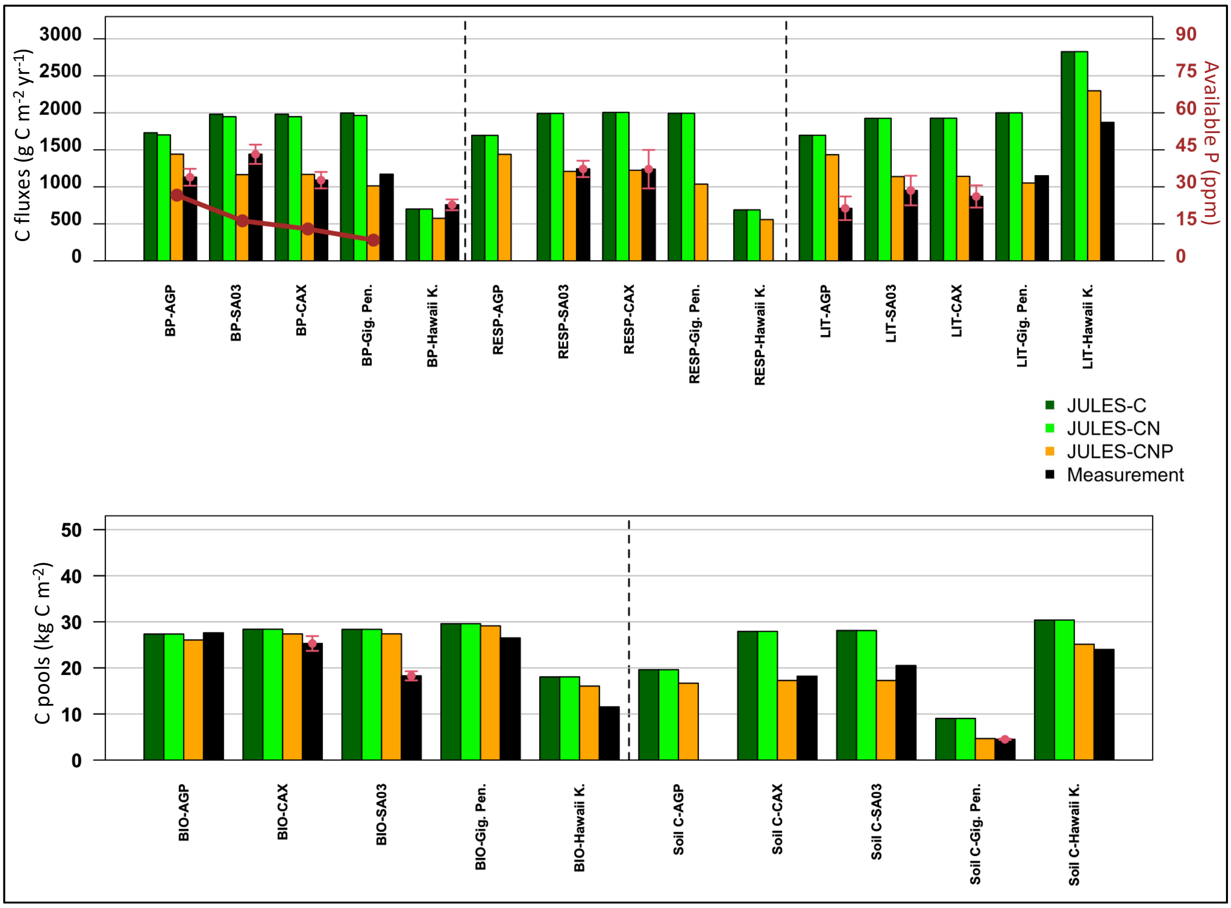

Figure. 4- Observed and simulated (JULES C, CN, CNP) C fluxes and pools (averaged measurements: red points, sd: red arrows) and available observed P (dark red points and lines (reported in ppm)) at test sites across the Amazon (AGP, SA03, CAX), Gigante Peninsula (Gig. Pen.) and Hawaii Kokee (Hawaii K.).

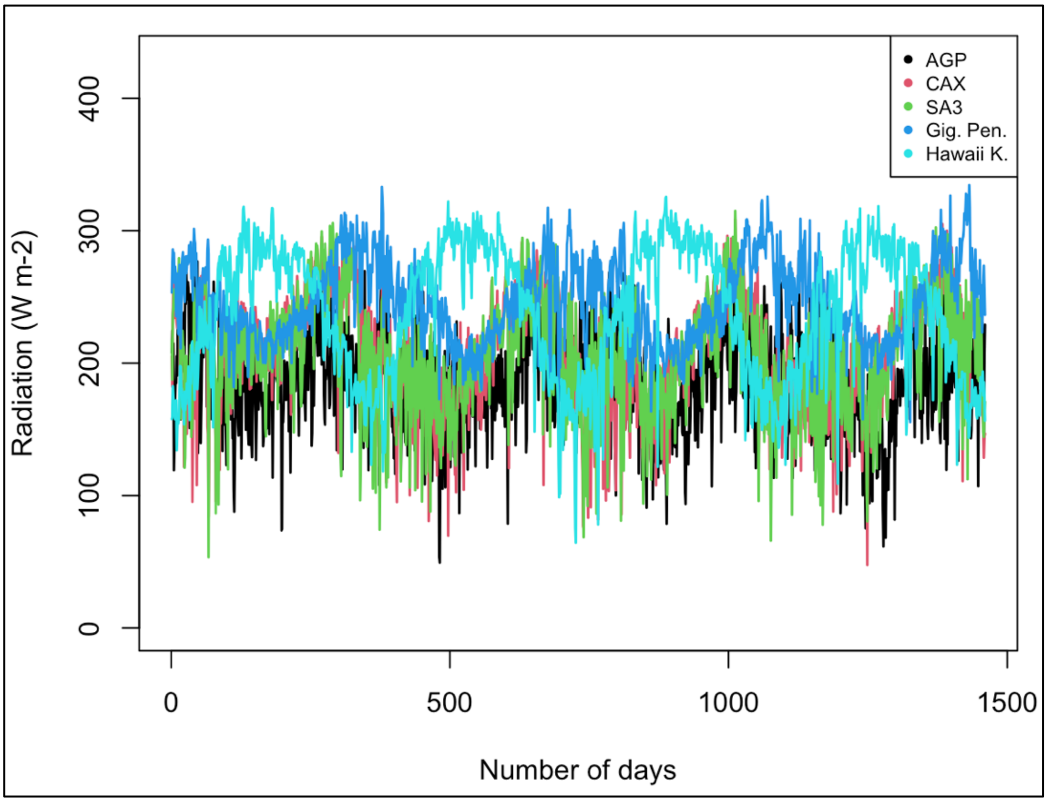

Figure. S8- Solar radiation at the extended test sites

RC1:

I thank the authors for the revision, but I fear the manuscript still suffers from two key flaws: the lack of novelty and a lack of model evaluation. Both flaws have been mentioned by more than one reviewer, who gave suggestions on how to address both points.

The authors argue that documenting the inclusion of a P cycle in LSM which hadn't had one before is novel enough. I have doubts that sufficient to qualify as 'substantial new concepts, ideas, or methods': The authors do not provide any new modelling concepts but have adapted published formulation, the single model application is a repetition of a multi-model exercise, the evaluation is insufficient to show the method is working reliably. Please see in the following more details about aspects.

We thank the reviewer for their feedback on the revised version. Please see below the answers to the main issues raised by the reviewer. Note, reviewer's comment in grey highlight and our responses in blue italic format, followed by the modified text in black colour.

Novelty:
The author's argumentation regarding novelty in the replies to Dr. Jiang and reviewer #3 are largely based on a misperception of the current state of science in this field:
There are several globally applicable CNP models (including land surface models) which emerged more than 10 years ago (e.g. CABLE(Wang et al 2010), JSBACH (Goll et al 2012), ELM/CLM (Yang et al 2014), ORCHIDEE (Goll et al. 2017)) and more in the pipeline (e.g. QUINCY (Thum et al 2019). JULES is merely another LSM which adds a P cycle. This is not a novelty of this paper.

As mentioned in the response to the editor, we have added few lines in the manuscript to explicitly say upfront that we are 'utilising state of the art already tested and implemented descriptions of P cycling in other land surface models (Wang, Houlton and Field, 2007; Zhu et al., 2016; Goll et al., 2017.' and that "JULES-CNP includes the representation of the P cycle in JULES version (vn5.5) and it is built on existing and well tested representations of P cycling in other global land surface models (Wang, Houlton and Field, 2007; Yang et al., 2014; Goll et al., 2017; Sun et al., 2021)."

There are several dedicated 'model-data nutrient cycling studies specifically for the Amazon forest with poor soils and limited P availability' in contrast to what the authors claim on page 3, namely: Goll et al 2018, Yang et al 2014, 2016, Fleischer et al 2019, Sun et al 2020 to name a few. This is not a novelty of this paper.
Perhaps overlooked by the reviewer, some of the data sets from our study listed in Table 4 used for either model evaluation/initialisation/parameterisation/evaluation or calibration are unpublished, from the Manaus region but still have not been used before in any modelling context.

The presentation of P processes is arguably not more detailed than in the first CNP models (Wang et al 2010, Goll et al 2012), and less detailed than in later models (processes missing here are: biochemical mineralisation, microbial dynamics, stoichiometric flexibility, atmospheric nutrient deposition, to name a few). There is no novel concept or process in this study.
Agreed and now explicitly upfront in the text as explained above.

Other studies also used site specific parameterization of their model when applying at site (e.g. Yang et al 2014, Fleischer et al 2016). This is not a novelty of this paper. Besides, the authors do on stringently use site specific data. E.g. They use weathering rates from Wang et al. 2010 which distinguish soils globally into three weathering classes based on very few data points. This led to unrealistic weathering rates (e.g. compare with Hartmann et al. 2014). Thus, most studies use either site-specific (optimised) rates (e.g. Yang et al 2014, Goll et al 2017) or use data-constrained global gridded weathering rates (Sun et al 2021) to resolve the large variation in weathering inputs (several orders of magnitude).
Other models have soil P structures which are based on measurable soil P pools, too (Yang et al 2014). This is not a novelty of this paper.
A side note: Wang et al 2018 is not a land surface model, but a biome-scale model designed to assimilate a wealth of observational data. The comparison is inappropriate.

We do not agree with the reviewer on this point. It was already mentioned in the revised version of the manuscript that despite the representation of the weathering processes in model, due to the simulation period,

we deactivated this process and instead prescribed a constant weathering release rate (similar to (Goll et al., 2017)), thus the argument the reviewer regarding the weathering process in JULES and the unrealistic estimation is not valid. This is defined in the line 548-549 as follows:
"Moreover, despite the initial representation of the parent material pool in JULES and its depletion through weathering (eq. 43), as the magnitude of changes in the occluded and parent material pools are insignificant over a short-term (20 years) simulation period (Vitousek et al., 1997), these two pools were prescribed using observations."

Please find the reaction to the novelty of model in the replies to editor (page 1 of this document).

Evaluation:
The heavy model calibration vs hardly any evaluation was criticised by all reviewers. The evaluation in the revised manuscript is still insufficient (i.e. a few C stocks and annual fluxes , and two soil P pools). The uncertainties in the observation does not allow us to distinguish if the C, CN, or CNP model is more realistic (Figure3). Given the large number of model parameters, it is not surprising the model is able to capture the two P pool stocks. The stocks are commonly used as target pools to optimise model parameters (e.g. Wang et al 2010).
No further attempt in model evaluation has been made. Dr Jiang and myself suggested using data from nutrient addition experiment(s) like e.g. from AFEX as the authors emphasised this experiment in their manuscript. But AFEX is not the only nutrient addition experiment (e.g the Gigante experiment in Panama) and other experiments have been used to evaluate models (e.g. Yang et al 2014, Goll et al 2017). Besides, there is other information which can be used to evaluate models (e.g. Sun et al 2021). The availability of data for model evaluation has to be considered when selecting the study site and in the model design, poor judgement regarding site selection is not an argument for insufficient evaluation. Land surface models are commonly criticised for being overparameterized (e.g. Prentice et al 2015); evaluation is key to avoid repeating past mistakes.

We have addressed this now with five extra sites, 3 from the Amazon across the west to east fertility gradient, one from west, one form central and one from easter amazon, one site from the Gigante Peninsula experiment and one from the Hawaii chronosequence.

Dr Jiang and myself expressed concerns that there is no data from the eCO2 experiment to evaluate the model, and it has been already simulated by a set of models. The authors claim 'there is a lot of value in knowing where our predictions lie compared to other models' with respect to the eCO2 experiment. However, they do not provide any further explanation of what the value would be or any example. In my opinion, there is little to none. I understand it might be interesting to a model developer to see how his models compare to others, but what is the scientific value?

Please find the response to the evaluation of model in the replies to editor (page 2-5 of this document).

RC2:

This manuscript is the revised version of a previous discussion paper. In my opinion, the authors have successfully addressed the comments of both myself and the other reviewers. The paper now includes a more complete and detailed model description and a comprehensive parameter analysis. It is now easier to follow and the results are more robust.

I would like to take a moment to address the comments of reviewer 3 regarding novelty. This is a model description study and its main purpose is to provide the reader with a very detailed description of the model development, be it novel or not. It is of great value to the community as it allows others to understand and reproduce the work done in particular models. It is also of great use to the authors themselves, as they can then publish the novel science without the need to describe the new model and the evaluation in such detail. Additionally, and perhaps most importantly, model development is a painstaking and time-consuming task that often gets very little formal recognition, especially for junior scientists and model description papers such as this are in part a formal recognition of developers' work.

We thank the reviewer for their positive feedback on the revised version, the suggested corrections and for defending the novelty and importance of this study, which was raised by reviewer 3. We have addressed the comments as described below. Note, reviewer's comment in grey highlight and our responses in blue italic format, followed by the modified text in black colour

A handful of minor comments:

L 144 Do the C:P and N:P ratios remain the same throughout the canopy?

Indeed, although the leaf N and P exponentially decreases through the canopy, the C:P and N:P ratios remain fixed throughout the canopy. We clarified this in the revised text as follows:

"Therefore, in JULES CNP in order to keep consistency with JULES C-CN, we also assume a multi-level canopy, and leaf N and P in exponentially decreases through the canopy (CanRadMod 6) (Clark *et al.*, 2011) while the C:P and N:P ratios remain the same."

l 176 Is the parent material pool a pool that can be depleted?

The parent material can be depleted using weathering rate over the parent material (eq. 43), however this operates over a much longer time scale than our study period (20 years) leading to insignificant changes in the pools. Therefore, these two pools are prescribed in model without consideration of the weathering process. This is further clarified in the second revision as follows (line 534-535):

"Moreover, despite the initial representation of the parent material pool in JULES and its depletion through weathering (eq. 43), as the magnitude of changes in the occluded and parent material pools are insignificant over a short-term (20 years) simulation period (Vitousek *et al.*, 1997), these two pools were prescribed using observations."

L 325 Check the edit here - the description varies spatially?

We corrected these lines as follows:

"Plant P uptake ($F_p{}^{up}$) varies spatially depending on the root uptake capacity (u^{max}) followed by Goll *et al.*, (2017). Therefore, in regions with limited P supply, the plant P uptake is limited to the u^{max} and consequently impacts the excess C and BP."

L 595 Each parameter was varied independently, yes?

Indeed, each parameter was tested independently. This is further clarified now as follows:

"To test the sensitivity of the P and C related processes to the model P parameters, six sets of simulations were conducted independently with modified plant C:P stoichiometry (Plant C:P: *SENS1*), P uptake scaling factor (K_P) (Kp: *SENS2*), inorganic (KP_sorb_in: *SENS3*) and organic (KP_sorb_or: *SENS4*) P adsorption coefficients ($K_{sorp-or}$, $K_{sorp-in}$), and maximum inorganic (KP_sorb_in_max: *SENS5*) and organic (KP_sorb_or_max: *SENS6*) sorbed P (K_{or-max}, K_{in-max})."

L 640 'The excess C flux is highly dependent on the plant P and the overall P availability to satisfy demand' aren't these the only things that excess C is dependent on?

Indeed, the reviewer is right. The excess depends only on the Plant P and inorganic P availability. We corrected these lines as follows:

"The excess C flux depends on the plant P and the overall P availability to satisfy demand (Table 5)."

L 891 Perhaps also worth mentioning here some papers that think p deposition will play a role, rather than just the one that doesn't , for example Gross, 2021https://doi.org/10.1111/nph.17344 or Van Langenhove 2020) https://doi.org/10.1007/s10533-020-00673-8 . (Please do not feel like I'm asking to cite these particular two, just any study that shows P deposition will contribute to plant and microbe available P)

Thank you for the suggested addition and references. We modified these lines as follows:

"Moreover, despite studies that show the possibility of P fixation as a source of available P for plants (Van Langenhove *et al.*, 2020; Gross *et al.*, 2021), due to the strong fixation of P in the soil (Aerts & Chapin, 2000; Goodale, Lajtha,Nadelhoffer, Boyer, & Jaworski, 2002), the P deposited is unlikely to be available to plants in the short term (de Vries et al., 2014), for this reason this version of JULES CNP did not include P deposition"

RC3:

I thank the authors for addressing my comments. I'm generally happy with most of their responses/revisions. One point that I would like the authors to further strengthen on is its novelty. Again, as I indicated in my previous review, I appreciate the extensive efforts went into this work; this is a great achievement to add P cycle into JULES. However, I'm not sure how to justify its novelty. Most of the P cycle processes included in this work are based on previously published literature, hence I would argue that their inclusion into a model is not the novelty per se. Maybe the authors should further strengthen why adding P into JULES is needed. Moreover, I can see that site-specific evaluation is valuable, and the authors have spent efforts to differentiate variables that are parameterized and those that are simulated, which addressed the concerns over over-parameterization. But for their predicted CO_2 responses, what can we learn given that there is no actual data to evaluate the predictions? The authors often refer to Fleischer et al. (2019), among other papers, to compare their simulation results. I understand that the simulated CO_2 responses of JULES fall into the broad spectrum of responses shown in Fleischer et al., which is great. But the value of Fleischer et al. is that they identified different model-based mechanisms to explain the predicted CO_2 responses. I would suggest the authors to make comparison in terms of the underlying mechanisms. The way the Discussion is written in its current form reads more like a comparison of numbers. It would be more useful for future model development's purpose to identify why the model predicts these numbers.

We thank reviewer for their positive comment.

Please find the answer to the novelty of the model and evaluation at further test sites in the replies to editor (page -1:3 of this document)