Reviewer 2

Many thanks for your constructive comments on our manuscript. Please find our responses below, with your original comments in regular text and our responses underneath in green:

'This manuscript investigates the effect of nutrient addition on grassland using a combination of process based modelling and observations from a manipulative experiment. The authors use the simple ecosystem model N14CP to simulate nutrient dynamics at two contrasting grassland sites in the UK and compare these results to data from a long term n and P addition experiment at the same sites. The study shows that P availability difference between the sites leads to differing limitation over time and differing effects of nutrient addition. The question of N and P limitation and co-limitation is very important and very topical, especially in the context of anthropogenic N deposition. Using process based models in conjunction with manipulative experiments is a very useful tool, not only for validating models but for advancing our understanding of ecosystem processes.'

Thank you, we are glad that you agree the research question is important and topical!

'One of the main issues of the manuscript in its current form is the way the observations are actually used to inform the model. As far as I can tell, the data is simply used to calibrate two parameters and then hardly ever mentioned again. The first problem here is the calibration itself: all experimental data is used at this step. This means that implicitly the model can represent observations from all treatments and lowers my confidence in the model's ability to predict responses to nutrient additions. I would suggest performing the calibration with the data from the control plots only, if there is sufficient data.'

Thank you for bringing this to our attention, you raise valid points that we should have clarified in-text. You are correct that the empirical data is used for calibration of $P_{CleaveMax}$ and P_{Weath0} but is seldom mentioned later in the manuscript. More references to the empirical data would certainly be useful for contextualising the results and will be added where relevant (such as figure 4 as you suggest later).

In reference to your point: 'The first problem here is the calibration itself: all experimental data is used at this step', we should highlight that we did exclude empirical biomass carbon data from the cost function to assess the model's ability to simulate empirical data that weren't provided. These outputs are provided in figure 2 panel a. As biomass C is the most responsive of the variables to nutrient additions (both in terms of rapidity and magnitude of response), we deemed this to be the most informative data to set aside for separate model testing.

However, we acknowledge that we are not sufficiently clear and are even contradictory in the methodology section for this where we state:

'The sum of the absolute errors between modelled and observed plant C and soil C, N and P data were scaled (to account for differing numbers of observations) and summed to provide an F value (Equation 1) as an overall measure of error across multiple observation variables.'

Whilst equation 1 does not contain information pertaining to biomass C, we imply in the text above that it is used and we only explicitly state that it isn't in supplementary section S1.1. We shall amend the text accordingly to make it clear that biomass C data was excluded and later used for blind

testing. These changes will be made towards the end of '2.3.3. Model parameters for the acidic and calcareous grasslands'

Regarding using just control data for simulations, we would agree with you that this approach would certainly be sensible and ordinarily deemed most appropriate for a model development study. However, we chose not to do this as the focus of this study is more explorative than it is model development. It was more important to us that the model could capture the potential differences in responses of each nutrient treatment, more so than how well it could simulate the sizes of C, N and P pools in each.

As our main aim was to explore how variation in P access mechanisms helps explain system responses to nutrient perturbations, we thought it better to include data relating to all nutrient treatments in the calibration. If we were to only include control plots, we may have missed the effects of P limitation being exacerbated by N addition captured by the empirical experiment. This is also why we chose to exclude empirical biomass data from all treatments from the initial calibration - to ensure that we could use the calibrated model to simulate responsive variables under all nutrient conditions.

It may be helpful if we provided some justification for this approach in the methodology section '2.3.3. Model parameters for the acidic and calcareous grasslands', so thank you again for raising this.

'The second data issue is the lack of model data comparison beyond figure 2. Specifically for figure 4 I wonder if it would be possible to add the observations to the plots rather than just referring to a supplementary table. It might even be useful to just show the experimental period, currently in figure insets, to better show how the model compares to the experimental results. This would increase our confidence in the model and build up the argument towards the predicted long term trends.'

We have previously explored adding the observations to figure 4 and decided to omit to focus on time series results instead. We could certainly add these back into the figure. We will also adjust this figure to provide greater focus on the experimental period, but still see the value in including time periods before this as it provides context for the magnitude of change and relative drivers of change (i.e. background N deposition Vs experimental additions).

'I also find that the paper lacks a discussion of the model's short-comings. While I understand the usefulness of simple models in allowing easier process attribution and avoiding over-parameterisation, N14CP is lacking a number of processes compared to state of the art vegetation models and this needs to be acknowledged. Most importantly, model NPP does not appear to include a response to increased CO2. This is particularly important for predictions of long-term nutrient limitation as elevated CO2 has been shown to increase plant nutrient limitation and I do not see how we can have any model predictions that do not take this into account.'

Thank you for your comment, and we will expand our discussion of model shortcomings – indeed, every model has short comings. Many models that include the effect of elevated CO2 have major short comings in their inclusion of nutrient effects. Hence, we have focused here on one of the few

models that does integrate carbon, nitrogen and phosphorus cycles for this study that centres on N-P limited grasslands. We agree that elevated CO2 does present an interesting problem – the research team leads such a MiniFACE experiment, to simulate ecosystem responses to concurrent eCO_2 and nutrient limitation. However, this study addresses an experiment under past/recent conditions where perturbations in N and P availability far outstrip natural changes in background CO2 concentration over the period, and hence we feel the use of this model is appropriate. This study provides a foundation for future work that could include CO2 effects. We will acknowledge this in the limitations discussion (see below).

'Other missing processes are less important, but would still need a paragraph in the discussion, especially the very simple plant pool structure and limited plant control on nutrient demand and uptake. It is even unclear to me if there is a belowground plant pool that would determine the N and P uptake or indeed if NPP scales with biomass. I do want to stress that I am not suggesting that authors modify their model or that the model is wrong, but that the assumptions in the model structure need to be highlighted and discussed.'

These are important processes to discuss in further detail that we shall amend accordingly, thank you for highlighting them. We shall add a separate sub section toward the end of the discussion to highlight these limitations.

In particular, we shall include detail on the potential effects of CO_2 enrichment on N and P availability, how these may be important and why they are currently omitted from N14CP. In addition, we shall discuss the simplicity of the plant pool structures and N14CP's simulation of plant control over nutrient uptake, and add clarification where required.

For example, and in reference to your comments, N14CP does have a belowground plant pool, represented as the root fraction of plant biomass. This was perhaps unclear as we didn't want to include too much repetition from the initial model development paper, which goes into great depth about model processes.

Plant biomass N and P content is somewhat determined by plants via different plant functional type transitions and changes in stoichiometry relative to specific plant end members which differ in their C:N ratios. These are responsive to environmental changes such as nutrient manipulation and / or N deposition. We aimed in this study to in part better represent plant control on P availability by varying $P_{CleaveMax}$ though this was not an explicit aim but may be worth mentioning.

Thanks again for raising these useful points.