Review of Winkler et al.:
Limitations of emergent constraints on multi-model projections: case study of constraining vegetation productivity with observed greening sensitivity

The authors explore the robustness of the "emerging constraint" (EC) method by using vegetation changes in the Northern High Latitudes as a case study. As the authors discuss, the EC method has gained increasing popularity and is being applied to a wide range of climate change studies, including reducing uncertainty in the carbon cycle. Overall the paper is well-written and easy to follow. The authors identify and analyse a number of caveats that may influence results from the EC method that are likely relevant to the wider community.

My main criticism is the use of LAI to predict GPP changes and stating that these two variables possess a strong causal relationship (indeed the authors state that the predictor and predictand should be causally related). Yes LAI and GPP are likely related, but there are many assumptions in models regarding how much GPP becomes NPP (i.e. how much GPP is respired) and how this carbon is then allocated into leaves, as opposed to other plant tissues. Furthermore, in the ESMs, allocation and respiration etc. can change with increased CO_2 forcing, influencing the GPP-LAI relationship over time. The authors should at the very least discuss the caveats of this approach and how this might affect their conclusions. I would also like the authors to consider in more detail what aspects of their findings might be specific to their case study (for example the idealised experiments where the effects of radiation and fertilisation effects are rather straightforward and increase GPP).

Specific comments:

P5 L21: Should this say 0.005 deg (500m) instead of 0.05 deg (5km)?

P5 L32: why averaged and not taking the max (to further reduce cloud contamination)?

P6 L26-28: Not very clear

P7 L7: Can you provide a few more details for "w" and how it was derived? Is it the time series for PC1?

P8 L15: Where do the vegetation classes come from?

P8 L17: I'm a little confused here. You talk about NHL but then go on to describe differences in tropical forests etc.

P11 L18: I admit I had to google the meaning of "Gedankenexperiment", perhaps a more common term is available?

P12 L9: do you show this anywhere?

P13 L14: Do these models simulate species composition changes?

P16 L1 onwards: these aren't really results presented in this study

Figure 2: I don't quite follow why only NHL was analysed when boreal, temperate forests and grasslands all show good agreement between AVHRR and MODIS (if this was the premise of the authors' choice)?