Referee 2

This is a neat study, carefully designed and carried out, with important implications for modelling afforested peatland systems. For the most part, the work is clearly described and well written. Some areas which need clarification are listed below. My main comment is that the significance of the finding that the peat is a net sink for carbon 30 years after being drained and afforested is very much under-played. This is contrary to expectations and current modelling assumptions, so merits more discussion.

Thank you for this positive feedback on our study. We agree that the net C sink in afforested peatlands is an important finding, and have expanded and in part re-written the discussion around this (also in response to another referee; see below).

Specific points:

L39-44. No, the reason why we are uncertain about the effects of drainage and afforestation is because it is logistically hard to measure. Even whether drainage actually causes a loss of carbon is based more on expectation rather than hard measurement.

We agree with this comment, and the preceding sentence (line XX-XX) states this quite clearly, We decided to remove the reference to numerical models, as our motivation is primarily to obtain data to establish drainage effects under forestry.

L97. The number of replicate plots is not given. Seems to be n = 4, repeated for each microform.

Added in: "the four paired"

L135. Was the moss layer conidered as "litter" and removed also? The text suggests there is only litter, peat and tree roots, but the moss layer looks non-negligible in the photo.

The moss layer was not removed. The site does have a sparse cover of mosses, we assume this only makes a small contribution to NPP, compared to the dense spruce/pine canopy and conifer needle input to surface litter.

L160. Soil moisture can be measured and expressed in many ways. Explain what is measured here - volumetric water content (m3/m3) by TDR method?

Added in: "measuring m^3/m^3 (volumetric water content)"

L173. For clarity, it would be best to give the equation for the final fitted model. "plot as a random effect" could be either an intercept or a grouping term on one or more of the other coefficients. The former I think, from Table 2.

Added: "(lmer(CO2 flux ~ (soil moisture *soil temperature * treatment * microform + treatment * microform * litter treatment) + (1|plot)))"

L179. The absolute values of AIC are very arbitrary, and there is no logic to saying that differences of less than 2 are meaningful. The relative values are meaningful, but there is no need to define such thresholds. The key thing is whether predictions differ substantially among these models - see point below.

The delta of less then 2 came from the referenced publication and we believe this is a useful way of comparing models. We appreciate that the referee has a different view on this.

L180+. Not sure why the weighting is mentioned, since it was not used. If there are notable differences between predictions from the different models, then using a weighted ensemble of model predictions would sensible. Bayesian model averaging would be even better. If, however, predictions are all rather similar, that justifies the approach of choosing the single best model (minimum AIC).

It is mentioned to show how confident we are in the top model, smallest AIC and highest weighting.

L168/L185. Fitted with nlme, but predicted with lme4? I think this is an error.

Indeed fitted and predicted with lme4, so changed in line 168.

L208. Only linear effects are considered here, but nonlinear effects are possible/expected, but harder to deal with and identify statistically. Can we get some justification for this?

The CO_2 data is log transformed, so we do consider non-linear effects.

L210. 40 % of variation was explained by the model, but this is presumably on the log scale. It needs pointing out that predictions are made in the original units, and all the uncertainties reinflate.

Added: “on the log scale, since predictions are made in the original units all uncertainties reinflate “

L215. Be explicit about the interpreation & units here - I think these are intercepts and multipliers for CO2 flux on the log scale (log(umol m-2 s-1) / deg C)?

We don’t think these need units, all numerical predictors were standardized to one standard deviation prior to analysis, so these fixed effects are used to interpret how big of an effect the particular predictor has with their standard error and p-value. We have added to the table heading the following to make this clearer: “All numerical predictors were standardized to one standard deviation prior to analysis.”

L215. There is no term for "Microform = Furrow". Maybe this is the interpretation of the interept?

Yes correct, this is the interpretation of the intercept.

L228. The negative interaction term just means that the T coefficient decreases with SM.

Figure 4 visualises what the negative interaction term means, which we have written in the text. We believe the text is correct.

L232-233. This is confusing, as it sounds like a separate step has been done. However, the whole rationale of fitting a model including soil moisture is precisely this - so that comparisons between treatments can be made, whilst accounting for differences in soil moisture.

This was indeed an extra step, and done for the model predictions only (so this is not about the model fit). By using the soil moisture data from just outside the plots when predicting the fluxes from the trenched plots, we took this artefact of the experiment away and were able to come up with a more accurate peat oxidation rate. This is explained in the Methods section: “The predictions were made over half-hourly measurements of soil moisture and soil temperature at 5 cm soil depth in all three microforms just outside the plots.” (line 194).

L235. Can you show the data as well as the fitted model?

This was plotted directly from the model output and we believe there is not an option to add in the data.

L238. For completeness, be explicit how heterotrophic fluxes are estimated - presumably as total - autotrophic.

Added: “The model prediction of heterotrophic respiration is calculated by subtracting the autotrophic flux from the total soil flux and includes emissions from decomposition of cut roots”

L243. Be explicit how the uncertainty on the annual sum is calculated. The error terms in Eqn 2 all add, and have to transformed from log to original units.

All error terms where indeed back transformed to original units, as well as the actual outputs and error terms were propagated. This has been added in the methods (Line 193 and 200).

L293. The table caption is very confusing. It reads as if this is the decay of dead roots itself, not the flux from the plot after the dead-root correction. Needs re-wording.

Changed to: “Corrected for dead root decay in trenched plots, heterotrophic (peat only; Fh) and autotrophic (Fa) fluxes (standard error in brackets) in g C m-2 y-1”

L378+. Of course there have to be codicils on the results, and it is the C balance over the lifespan of the forest that matters. However, the expectation in most modelling work is that drainage and afforestation causes oxidation of the peat at a rate of 50 to 300 g C m-2 y-1 (e.g. Cannell 1993). This is offset in the first few rotations by the increasing tree biomass and litter, but ultimately, the ongoing long-term degradation of the peat becomes the dominant term, and the system becomes a net carbon source after 1-5 rotations (depending on the assumed peat oxidation rate). If this study is in fact showing that the peat is a net sink of 17 to 124 g C m-2 y-1 after 30 years, this is surely the stand-out result. Worth some more discussion at least.

We have expanded and in part re-written the discussion around this (also in response to another referee; see below).