Authors response to referee comments

O'Shea, S. J., Allen, G., Gallagher, M. W., Bower, K., Illingworth, S. M., Muller, J. B. A., Jones, B., Percival, C. J., Bauguitte, S. J-B., Cain, M., Warwick, N., Quiquet, A., Skiba, U., Drewer, J., Dinsmore, K., Nisbet, E. G., Lowry, D., Fisher, R. E., France, J. L., Aurela, M., Lohila, A., Hayman, G., George, C., Clark, D., Manning, A. J., Friend, A. D., and Pyle, J.: Methane and carbon dioxide fluxes and their regional scalability for the European Arctic wetlands during the MAMM project in summer 2012, Atmos. Chem. Phys. Discuss., 14, 8455-8494, doi:10.5194/acpd-14-8455-2014, 2014.

We are very grateful to both referees for their insightful and helpful comments about our manuscript and for recognising the important scientific utility of the study. We have made use of those comments to improve the manuscript in our revision submitted with the response below. We particularly welcome the comments on restructuring of the manuscript to present a more coherent narrative. We now address the comments individually. For clarity, referees' comments are coloured red and the responses are coloured black.

Referee #1

O'Shea et al. investigate northern Fennoscandian wetland methane (and to a lesser degree CO_2) flux using a combination of chamber, eddy covariance and aircraft measurements and compare these against two large-scale models. The authors suggest that wetland area is underrepresented in the models, which results in an underestimation of flux magnitude. Aspects of the analysis are interesting, but the investigation is carried out at too many spatial scales, from chamber fluxes (on the order of centimeters and seconds) to model runs on the scale of decades and hundreds of kilometers. These multiple scale mismatches mostly confuse the message and obfuscate any simple apples-to-apples type comparisons. It follows then that much of the analysis is what I often call a 'bunch of stuff' (analyses that are not fully related to one another) with the authors questioning their own approach for example on page 8471 line 9. An on page 8474 line 18. The analysis was carried out competently in many regards, but in my opnion it would help the reader if the focus was simplified to describe the aircraft measurements in more detail and leave the chamber, eddy covariance, and modeling work - most of which comes from other analyses - to a comparison in the discussion. Right now thee paper is trying to do three things, instead of one thing well.

Addressing the temporal and spatial scale mismatches between different measurement techniques is a key novelty of this paper and a very important issue. It is precisely the potential and rationale for scalability that is important to science (and this paper) and our study begins to examine how this might reasonably be done and what some of the limitations might be. We believe studies that link datasets in this way are just as important and useful as those that treat datasets in isolation. All flux measurements have their own specific advantages and disadvantages, which may be related to their spatiotemporal coverage or to do with assumptions made in their calculation. Combining different approaches helps to resolve the true magnitude and variability of fluxes. While comparison between techniques can highlight biases and help to verify whether assumptions are sound. This paper provides a novel attempt to link different approaches together, it shows that small scale observations can be up-scaled to represent regional scales as long as there is sufficient knowledge about the land type in a region and the up-scaling method can be tested using a suitable constraint. This has allowed a significant low bias to be identified in two commonly-used process-based land surface models, which would not be possible with a single technique alone. The referee points out that we question our own approach on two occasions. Surely this should be commended; assessing the challenges, uncertainties and limitations of our work is an important part of the scientific method. We find it entirely proper to reason (and question) our approach in this regard such that future work by others who may seek to link spatial and temporal scales may learn and improve on it where possible.

We would like to thank the reviewer for their advice regarding the restructuring of the paper. We agree that this is a measurement led study and that the aircraft measurements should feature prominently with the comparisons as a more focussed discussion to follow. To this end, we have added an additional section that focuses only on the airborne measurements, which includes:

1) A subsection describing the mass balance approach and it implicit assumptions.
2) A subsection describing the sampling and study area.
3) A subsection describing the meteorology during the flight.
4) Back trajectory plots showing the synoptic airflow and airmass history.

Further details have been added to the manuscript regarding the airborne flux calculation and it corresponding uncertainty, such as estimates of the PBL variability during the flight. As the reviewer suggests, all comparisons with the chamber, eddy covariance, and process models are kept separate in the discussion section.

Specific comments:

’The Fennoscandian wetlands’ in the abstract probably doesn’t encompass their entire extent. (Reading the paper confirms that this is the case.)

Agreed – we have replaced the word “wetlands” with the word “landscape” and the latitudes/longitudes of the study.

Page 8458 line 9 is vague: please describe these feedbacks.

This has now been clarified to read:

“Palaeo-records indicate that strong positive feedbacks exist between climate and greenhouse gas emissions in the region, whereby warming causes enhanced emissions that in turn leads to further warming (Walter et al., 2007; Nisbet and Chappellaz, 2009)”

The parenthetical comment on page 8458 line 17 is a bit distracting, and is highly quantitative. How are these approaches poorly constrained? What critical uncertainties remain? Use this as a motivation for the rpesent research; at the moment it sounds like a poorly justified and vague swipe at previous efforts.

Agreed- the parenthetical comment has been removed from the revised manuscript.

The statement at the bottom of the page regarding methane consumption is well-put. Note also the recent manuscript by Parmentier et al. http://www.biogeosciences.net/8/1267/2011/bg-8-1267-2011.html

Thank you. This reference has now been added to the revised manuscript

On page 8459, Land surface models run at far more resolutions than just 0.5 degrees.

This is referring to global applications of land surface models. The text has now been clarified to read:

“Currently, there is a lack of flux measurements at the same spatial scale as the resolution of global land surface models (typically 0.5°)”

Far too many abbreviations in section 2.1. It’s just as easy to write - and easier to read - ’whole air samples’ rather than WAS.

We agree. The abbreviation “WAS” has been replaced with “whole air samples” throughout the text. All unnecessary abbreviations have been removed.

In section 2.2, why was the Webb et al. correction applied to the closed path IRGA measured fluxes? It needen’t be, unless it was decided that the tube length is too short to fully attenuate the effects of pressure and temperature fluctuations. In this case, often a partial Webb et al. filter is applied.

The reviewer is correct. The air density correction was only made for latent fluxes and not sensible heat fluxes. The manuscript has been changed to now read:

"An air density correction related to the latent heat fluxes was conducted according to Webb et al. (1980)."

I question the use of asymptotic fits to the chamber data, unless the model is to better-fit the linear portion of the concentration/time curve. Saturating concentration/time curves often indicate that the effects of the chamber are obscuring the ability to measure flux. Please describe this section in more detail.

The approach used is based on our previously published paper Levy et al. (2010) in which multiple models are fitted to the data and the most appropriate model chosen according to a number of criteria (e.g. the model measurement correlation coefficient). In cases where the curves are asymptotic, the model accounts appropriately for any curvature and estimates the flux at time zero. For a complete description of the method and its efficacy, see:

Levy P.E., Gray, A., Leeson, S.R., Gaiawyn, J., Kelly, M.P.C., Cooper, M.D.A., Dinsmore, K.J., Jones, S.K. & Sheppard, L.J.: Quantification of uncertainty in trace gas fluxes measured by the static chamber method. *European Journal of Soil Science*, **62,** 811-821, 2011.

The reference Levy et al., (2010) and further description as above, has now been added to the revised manuscript.

I know that the purpose of this manuscript is to test an upscaling approach and not explore different model formulations, but in equation 1 I would doubt that a carbon substrate parameter is needed. Methane efflux is probably not ever carbon limited in these systems. Also, is there any reason to believe that the default parameter set should be changed to better-fit observations?

JULES and Hybrid8 aim at simulating global methane emissions, not specifically northern wetlands. The amount and quality of substrate is one of the main drivers for methanogenesis. This may be not limiting for northern wetlands, but it allows discrimination between different soils around the world. Note that some models use primary productivity (NPP) instead of soil organic matter (SOM) for a similar objective (Melton et al. 2013). For a regional application the k(CH4) tuning factor could be adjusted to better fit the observations instead of modifying the substrate. However, as the reviewer notes this would be beyond the scope of this current study (but will be addressed in future ones).

Page 8464 line 24 and onward sounds like an advertisement for this model rather than a succinct and technical description of its capabilities. How is the canopy representation sophisticated and canopy conductance realistic?

Line 24 has been reworded and all non-scientific language has been removed. It now reads:

“This model contains a canopy representation that has a mechanistic canopy conductance response to various environmental factors (light, temperature, humidity, CO_2 and canopy height), which has been tested and calibrated using eddy covariance flux measurements (Friend and Kiang, 2005).”

Further regarding the model assumptions, CH4 flux is controlled by diffusion, advection (e.g. through plant aerenchyma), and ebullition by bubbles. What was done to account for the (potential) impacts of the other two transport processes?

The standard versions of JULES and Hybrid8 do not consider these different pathways. The plant mediated pathway is important for the CH4 budget. In this case, a significant amount (~50%) of the methane going through plant aerenchyma may be oxidised before reaching the atmosphere. However, this process is strongly ecosystem dependent (Bridgham et al., 2013) and difficult to integrate in global scale models. Plants with aerenchyma and their efficiency would need to be identified, which is not well constrained. Similarly ebullition is likely to be highly variable in space and time, difficult to constrain and difficult to represent in a global model.

However, this is an important area where future studies using the MAMM dataset (chambers, primarily) may provide useful insight into the importance of these processes and how different vegetation types could be introduced into these models. The current versions of these models both assume that temperature and water table depth are the main drivers of emissions. The recent paper by Yvon-Durocher et al. (2014) shows a consistent temperature response curve across ecosystems, which helps to justify the simple approach followed by JULES and Hybrid8.

I found the error propagation approach to be sound. It assuaged many detailed concerns that admittedly would be a bit tough to measure, like PBL height and entrainment dynamics.

We thank the reviewer for recognising this. This is very important to discern what we can (and can't) interpret from the data.

In section 3.3, the eddy covariance footprint depends on measurement height, sensible heat flux, and wind statistics. Sometimes it is in the 100 m to 1000 m range. What is the representative footprint dimension of the study site during the campaign?

The purpose of this sentence is just to comment that eddy covariance fluxes from the mast and airborne measurements operate in different spatial scales. As the reviewer notes the footprint varies on stability, however, whether the mean eddy footprint is 150 or 300 m it is still significantly smaller than the airborne measurements.

It is interesting to show that the models dramatically underestimate methane flux, but 30+ years of data (that don't even encompass the measurement domain) is unnecessary to do so. The interesting part of this analysis is that wetland extent appears to be underestimated, but the authors don't attempt to quantify by how much except to run the models under the assumption that all pixels are wetland, which is only partially realistic (all pixels almost certainly contain some wetland).

This comparison is intended as a very broad and mainly qualitative statistical comparison which highlights a systematic and consistently-observed low bias in the models with respect to wetland flux. Since the submission of the paper, we have now repeated the JULES run using the updated and extended (to December 2012) version of the NCEP-CRU meteorological dataset. This has not changed the basic result that the land surface models under predicted the methane emissions in this region. We have also included the results from using the same months from the complete model run (1980-2012). The results show that 2012 was not an anomalous year; the emission fluxes were slightly higher. Furthermore, by assuming all pixels are wetland in the models, we have attempted to diagnose what the model fluxes would be in the absence of (known) systematic problems associated with wetland area extent, thereby providing a maximal model flux climatology (which now just encompasses these measurements). We would prefer not to quote a fixed quantitative bias for the exact reasons that the reviewer has already noted, i.e. this is not a direct comparison, rather it is qualitative and highlights a systematic bias. Deconvolving errors in the wetland extent from those in the temperature term and soil carbon content would not be possible with the available dataset. However, as the reviewer notes the purpose of this paper is not to test different model formulations, rather it is to test an up-scaling approach, which has highlighted a significant low bias for these models in this region.

In Figure 3, don't use red and green at the same time unless necessary. Here it is unnecessary.

We agree. Red and green are no longer used in the same plot.

A higher quality Figure 6 is needed.

This appears to be a conversion problem with the figure into ACPD. The figure is an encapsulated post-script file of high resolution and looks fine locally. We will discuss this with the type-setter on potential publication of a revised manuscript.

Referee #2

The manuscript presents flux estimates for methane and carbon dioxide derived from different methods on different scales. The core novelty of this study is the determination of regional-scale greenhouse gas fluxes based on airborne measurements, which are then compared to eddy tower fluxes and chamber fluxes within the flux footprint, and which are then used to assess the skill of two land surface models, i.e. JULES and Hybrid8 for a study area Northern Fennoscandia.

As the airborne fluxes are the main new feature of this study and they are not published elsewhere before, it is particularly important that the methodology to obtain these fluxes is well documented, in order to allow an assessment of their representativeness and reproducibility. However, the description of the flight design is vague and sometimes unclear. The authors write (p8463 l16) they performed transect flights at an altitude range of 100 m to 2000 m. What exactly were the flight levels for the different transects and how constant was the flight altitude during one transect? Later in the manuscript (p8468 l15), they write that the range for the transect flights was 70 to 1287 m and 103 to 1382 respectively. If I interpret this information correctly, this is a very unorthodox flight experiment design, as the aircraft apparently changed its altitude and its east-west position simultaneously. Hence, it is not possible to distinguish whether the observed changes are due to the movement in the horizontal direction or in the vertical direction.

Please first see our response to reviewer 1 on the restructuring of the manuscript to focus more on the measurements.

We agree with the reviewer that the description of aircrafts altitude during the East and West transects was not clear enough. The altitude range 100 m to 2000m was incorrect and has been removed. We have now added an additional plot (Fig. 2c), which explicitly shows the aircraft's altitude when the measurements used in the flux calculation were collected. Due to the necessity of using discrete point measurements sampling needs to be optimised to characterise both the horizontal and vertical distribution of CH_4. During this work sampling was performed, where possible (due to air traffic restrictions) to provide both pieces of information. Only performing transects at one level, as the reviewer suggests, does not provide information on vertical mixing and as a consequence might bias the flux calculation. As can be seen from Fig 2, the horizontal CO_2 and CH_4 gradients are not strongly related to changes in the vertical.

The assumption needs to be made that the species of interest is vertically well mixed throughout the PBL. We have investigated this assumption using vertical profiles (Fig. 3), dispersion modelling (Sect. 4.2) and the calculation of a mean mixing time of 57 minutes, which has been determined using the convective velocity scale (Stull et al., 1988; Karion et al., 2013) (Note- this has changed from 80 minutes in the original manuscript since an error has been identified in its calculation, however this does not change the conclusions.)

In addition, there is a possibility that the observed changes are due to nonstationarity of the PBL. Maybe I am misinterpreting the authors' description but then it needs to be clarified in the manuscript.

We have added further information regarding the determination of PBL height in a new "Meteorological summary" (Section 3.3):

"Deep vertical profiles of potential temperature (derived here from in situ measurements of pressure and temperature) from the FAAM BAe-146, performed over Sodankylä (Fig. 3) at 1:00 GMT and 15:00 GMT and from the two dropsondes released, show a clear capping inversion was present over the area during the flight (Fig. 3). Over the run in question, the surface topography was very flat (400-500 m above mean sea level) and the infrared emissivity varied little (~0.98, see Allen et al., 2014). Therefore, in the absence of significant synoptic meteorological changes, which were not observed in reanalyses for the area, it is expected that the PBL depth was relatively uniform over the time and scale of the sampling in question.”

Further details have been included in Section 4.1 concerning the uncertainty estimates resulting from variation in the PBL height throughout the flight:

“In order to estimate the uncertainty in the determination of the PBL height we use a simple PBL growth model to estimate the change that could reasonably be expected in the intervening period between the nearest vertical profile and the completion of the longitudinal transect used in the flux calculation (approximately 1 hour). The change in PBL height, Δz, over the time period Δt can be estimated using Eq. 4 (Stull et al., 1988; Cambaliza et al., 2014),

$$\Delta z = \left(\frac{2\Delta t \overline{w'\theta'}}{\gamma} \right)^{0.5} \tag{4}$$

where γ is the adiabatic lapse rate and $\overline{w'\theta'}$ is the surface sensible heat flux, which was measured in Sodankylä. Using Eq. 4 changes in the PBL depth are estimated to be of the order 200 m within 1 hour, which we use as an estimate of the uncertainty in the PBL height during the transects.”

The mass-balance approach relies on a series of assumptions and these need to be verified through measurements whenever possible. These are well-mixed conditions (vertical profiles), stationarity (always flights along the same track in both directions, preferably around midday), zero entrainment flux (stable boundary layer height, check by several soundings in between), and horizontally homogenous wind field (flights at several levels). A much more detailed description of the flight experiment design is necessary to be able assess whether the airborne measurement strategy is valid. As it is now, too many questions remain unanswered. In general, this study lacks focus and therefore I would recommend to focus much more on the description, analysis and interpretation of the airborne measurements. In my opinion, major revisions are required and the revised manuscript will need another round of reviews.

As mentioned in our response to reviewer 1, the assumptions involved in the mass budget calculation have been made more explicit in the revised manuscript as has the estimates of the uncertainties resulting from these assumptions. In section 3.1, we have provided a full description of the mass balancing method and its implicit assumptions.

We have included a rigorous uncertainty analysis using the range of unknowns that might impact our assumptions. In addition, we have tested the derived flux using the NAME dispersion model. Reviewer 1 has praised this approach and the rigour of our uncertainty analysis. However, we agree that further description of the flight design is necessary (see response above) and we have now included further description to help the reader understand the sampling (and its limitations as raised by the reviewer)

Just one technicality: Figures 1 and 6 are much too small.

Similar to our response to reviewer 1 this appears to be a conversion problem with the figure into ACPD. We will discuss this with the type-setter on potential publication of a revised manuscript.

References

Bridgham, S. D., et al.: Methane emissions from wetlands: biogeochemical, microbial, and modeling perspectives from local to global scales, Glob. Change Biol., 19, 1325–1346, 2013.

Karion, A., et al.: Methane emissions estimate from airborne measurements over a western United States natural gas field, Geophys. Res. Lett., 40, 1-5, doi:10.1002/grl.50811, 2013.

Melton, J. R., et al.: Present state of global wetland extent and wetland methane modelling: conclusions from a model inter-comparison project (WETCHIMP), Biogeosciences, 10, 753-788, doi:10.5194/bg-10-753-2013, 2013.

Stull, R. B.: An Introduction to Boundary Layer Meteorology, Kluwer Academic Publishers, Dordrecht, 670 pp., 1988.

Yvon-Durocher, G., et al.: Methane fluxes show consistent temperature dependence across microbial to ecosystem scales, Nature, **507**, 488–491, 2014.