Reviewer 1 Comments

Comment) The authors address an important knowledge gap of ecosystem-climate interactions; the variation of understory carbon balance during a growing season and its interactions with nitrogen status. In general, the manuscript reads very well, almost so well that omissions are easily overseen. An important omission is that no information about plant development at the time of measurements is given at the time of the measurements. One would expect a net uptake of C at the time of leaf development.

Response) Thank you for highlighting this omission. We have some data about plant development during the year, in the form of plant greenness calculated for each location and time period from digital photographs, which we now include in the manuscript with additional text, graphics and analyses (Table 2, Figure 2, Lines 193-200, 258-260, 267-269, 296-298). This method of measuring greenness with digital photographs has been used successfully in several published papers (Ahrends et al. 2009, Migliavacca et al. 2011, Richardson et al. 2007, Westergaard-Nielsen. et al. 2013), from a diverse range of vegetation types and ecosystems, where it has been found to correlate closely with the most common remotely-sensed index of vegetation density (normalized difference vegetation index) and GPP. We had omitted this data because we thought that stem density provided a similar, more direct proxy for vegetation density, but we now recognize the additional value of measuring temporal change in vegetation. In addition, in Figure 2 we now also summarize temporal change in soil temperature and moisture.

Ahrends, H.E. et al. 2009. Tree phenology and carbon dioxide fluxes: use of digital photography for process-based interpretation at the ecosystem scale. Climate Research 39:261-274.

Migliavacca, M. et al. 2011. Using digital repeat photography and eddy covariance data to model grassland phenology and photosynthetic CO_2 uptake. Agricultural and Forest Meteorology 151:1325–1337.

Richardson, A. et al. 2007. Use of digital webcam images to track spring green-up in a deciduous broadleaf forest. Oecologia 152:323-334.

Westergaard-Nielsen, A. et al. 2013. Camera derived vegetation greenness index as proxy for gross primary production in a low Arctic wetland area, ISPRS Journal of Photogrammetry and Remote Sensing 86:89-99

Comment) Furthermore, information about the exact dates for the measurements P14097, L19-20 are missing, this is relevant information as the C balance in the first halve of June might differ from that in the end of June. V. Myrtillus is known for its green twigs and stems so that it can start photosynthesis as early as possible in the season. It can therefore, with the available information, not be ruled out the authors missed to measure an essential period with high GPP of the understory carbon balance.

Response) We now include this information in the methods section (Lines 145-147). Clearly, the experimental design of the study, with three snapshot measurements through the growing season, provides only coarse information about the temporal development of forest floor CO_2 fluxes, and we acknowledge that we likely missed the main period of early season leaf flush (Lines 300-301). Measurements with automated chambers are ideally suited for continuous measurements, but are difficult to replicate over space, which is essential to derive a robust estimate for that ecosystem, and if you want to assess CO_2 fluxes across a range of experimental treatments, as we

do. Therefore, each approach has its own strength and weaknesses but we believe that the current study provides useful information about the magnitude and drivers of spatial variation in forest floor CO_2 fluxes, and highlights broad-scale temporal trends which can be investigated in further detail with other, complementary approaches.

Comment) Concerning the conclusions, I find the importance attached to understory for respiration not well grounded. Measured respiration is the sum of above-ground understory respiration, root respiration of understory and trees, and soil respiration due to decomposition. The soil temperature was by far the best explaining environmental factor, indicating that explanation for total respiration should be sought in the soil, due to root or decomposition processes. Reviewer III has a good point in bringing up tree root respiration. I wish to add that the contribution of decomposition processes need to be considered. The nitrogen fertilization started a five years before the measurements. This time period is long enough to expect an impact of the N fertilization on the carbon and nitrogen mineralization rates, but might be too short for significant effects on the thickness of the organic layer. Needles, the main litter source, can have life-span around five years in a maturing boreal forest.

Response) We acknowledge that some of our use of the term "understorey" in the manuscript was misleading because we actually measure the contributions not only of understorey vegetation, but also tree roots and soil, as you point out. We did acknowledge this in the original manuscript (Page 14098, Lines 11-12), but nevertheless, to clarify this issue, we have replaced "understorey" with "forest floor" throughout the manuscript. This term gives a clearer description of what we actually measured, and has been used in a large number of published papers making similar measurements to our study (e.g.: Drewitt et al. 2002; Gaumont-Guay. et al. 2014; Jassal. et al. 2005; Kelliher. 1999; Morén & Lindroth 2000; Swanson & Lawrence 2001). In one portion of the analysis, calculation of the "plant carbon surplus potential" (Figure 3), we still explicitly refer to understorey vegetation because in this case we utilize respiration data from Hasselquist et al. (2012) which specifically isolated the contribution of understorey vegetation to forest floor CO_2 efflux. This was already stated, but is now clarified further, in the introduction, methods, and results sections, and in the Figure 3 legend (Lines 106-109, 219-221, 278-281, 603-605). We now take care to refer to the fluxes measured by Hasselquist et al. (2012) as R_{uv} (for understorey vegetation) rather than R_a (for autotrophic) throughout the manuscript, which we acknowledge had been a source of confusion because tree roots are also a source of autotrophic respiration. We also now include a substantial extra section in the methods section summarizing the methodology of the Hasselquist et al. (2012) study and describing how their data were used for comparison with our data (Lines 219-233).

We agree that there are good reasons to expect that soil heterotroph and tree root contributions to forest floor CO_2 efflux could change under fertilization. We discuss N impacts on decomposition and respiration, with references, in the introduction (Lines 77-93). However, our measurements are not designed to tackle this issue, and it has already been largely addressed with the analysis of Hasselquist et al (2012) that was specifically designed to partition forest floor respiration. We believe that the novel contribution of our study is in quantifying the magnitude and controls on both uptake and release of CO_2 by the forest floor as a whole (incorporating ground vegetation, tree roots and soil). This data is of considerable scientific value because it can be directly compared with other integrated fluxes (whole forest net ecosystem exchange and respiration) with minimum error. A large number of published studies highlight the value of these types of integrated measurements (e.g.: Drewitt et al. 2002;

Gaumont-Guay. et al. 2014; Jassal. et al. 2005; Kelliher. 1999; Morén & Lindroth 2000; Swanson & Lawrence 2001).

Drewitt, G. B. et al. 2002. *Measuring forest floor CO_2 fluxes in a Douglas-fir forest. Agricultural and Forest Meteorology 110:299-317.*
Gaumont-Guay, D. et al. 2014. *Eight years of forest-floor CO_2 exchange in a boreal black spruce forest: Spatial integration and long-term temporal trends. Agricultural and Forest Meteorology 184: 25-35.*
Jassal, R. et al. 2005. *Relationship between soil CO_2 concentrations and forest-floor CO_2 effluxes. Agricultural and Forest Meteorology 130:176-192.*
Kelliher, F.M. 1999. *Carbon dioxide efflux density from the floor of a central Siberian pine forest. Agricultural and Forest Meteorology 94:217-232.*
Morén, A. & Lindroth, A. 2000. *CO_2 exchange at the floor of a boreal forest. Agricultural and Forest Meteorology 101:1–14*
Swanson, R. V. & Lawrence, B. 2001. *Environmental regulation of carbon dioxide exchange at the forest floor in a boreal black spruce ecosystem. Agricultural and Forest Meteorology 108:165-181.*

Comment) The title needs to be adjusted; change 'the' growing season into 'a' growing season or rather a summer as only one summer was studied. The understory carbon balance might vary with the forest development stage (see for example Hedvall P-O et al. 2013) and weather conditions. (Similar changes needed in abstract L6, L16 and conclusions L19). P14095, L12-13 add 'and cold-temperate' to boreal (Gärdenäs,2000 is a study in the cold-temperate zone)

Response) We agree that we need to emphasize that we measure over only one growing season, and therefore that our measurements may reflect only patterns over that particular season, not necessarily all growing seasons in general. We believe that the correct way to do this is to refer to "the" growing season rather than "a" growing season because "the" (as the definite article in English), implies that there is a particular growing season that we are referring to whereas "a" (as one of the indefinite articles in English, the other being "an") growing season implies that we are referring to any or all growing seasons in general.

We are not clear where it would be appropriate to insert the Hedvall et al. (2013) reference. If you could direct us to the specific place where this reference would be useful, we would be happy to insert it. We have changed text in Page 14095, Lines 12-13 as suggested, thank you for pointing our this error (Lines 70-71).

Comment) P14098, L25 and Figure 1. Belongs to results and need to be reformulated. Figure 1, an example of autocorrelation? First you derive an exponential model of NEE based on NEE measured and then you compare NEEmodelled with NEEmeasured (Figure1). I suggest to divide NEE measured data in two halves, use one half to derive your model and the other half to test your model.

Response) We now present these data in the results section (Lines 237-240), and have clarified the exact purpose and limitations of the modeling approach taken in this study (Lines 174-186), which is well established and widely used (e.g.: Cahoon et al. 2012; Kolari et al. 2006; Street et al. 2007; Upek et al. 2008; Williams et al. 2006; Shaver et al. 2013). See for example, Figure 1 in Shaver et al. (2013) which is essentially exactly the same as our Figure 1 but presented for multiple sites. Their model fitting method is also very similar, they also minimize the root mean

squared error between observed and predicted NEE values by varying model parameters with Excel solver until the optimal combination of parameters is found *for that particular location and time period.*

We emphasize that the model is used only to predict NEE for a uniform light level (600 μmol m^{-2} s^{-1}) for each location based upon the measured light response curve specific to that location and time period. Hence, we developed 153 unique model parameter sets (51 locations × 3 time periods). The models are not intended to be used, and are not used in the study, to predict NEE beyond the immediate conditions over which they have been parameterized, such as across the season or at other locations. Therefore, we do not think it would be appropriate or informative for our analysis to use half of the data to fit models and half to test the models because this would be validating the models for a purpose for which they were never used, or intended to be used.

We are not clear how measured and modeled NEE could be an example of autocorrelation because measured NEE and the three parameters which combine to predict NEE are not intrinsically related to each other. As such, using our modeling approach it is certainly possible for measured and modeled values to not be closely related to each other. In our case, there is an excellent fit because: (1) NEE and GPP follow a very well established and highly predictable response to light, (2) Generally the more parameters included in a model which are allowed to vary, the closer the fitted value will be to the target (observed) value.

Cahoon, S. M. P. et al. 2012. *Large herbivores limit CO_2 uptake and suppress carbon cycle responses to warming in West Greenland. Global Change Biology 18:469-479*
Kolari, P. et al. 2006. *Forest floor vegetation plays an important role in photosynthetic production of boreal forests, Forest Ecology Management 221:241-248.*
Shaver, G. et al. 2013. *Pan-Arctic modeling of net ecosystem exchange of CO_2. Philosophical Transactions of the Royal Society of London – Biological Sciences 368, 20120485, doi:10.1098/rstb.2012.048.*
Street, L. E. et al. 2007. *What is the relationship between changes in canopy leaf area and changes in photosynthetic CO_2 flux in arctic ecosystems?. Journal of Ecology 95:139-150.*
Upek, B. et al. 2008. *Forest floor versus ecosystem CO_2 exchange along boreal ecotone between upland forest and lowland mire. Tellus B, 60: 153–166.*
Williams, M. et al. 2006. *Identifying differences in carbon exchange among Arctic ecosystem types. Ecosystems 9:288-304.*

Comment) P14099, L2-3. Which kind of probes were used and for how long time was the soil temperature and moisture measured. Did the measurements cover the variation of a whole day?

Response) We now describe the types of probes used, and clarify that they were used for instantaneous, rather than continuous, measurements of soil moisture and temperature (Lines 187-191). Therefore, the soil temperature and moisture data represent conditions exactly at the time of CO_2 measurements. For our purposes, to correlate CO_2 fluxes with abiotic and biotic factors, we believe that this level of sampling is sufficient. We now also clarify that the CO_2 and supplementary measurements were performed between 9:00 and 16:00 on the sampling days (Line 172).

Comment) P14100 L21, add 'soil' This lag between R and GPP meant that the understory and soil

Response) We have altered this sentence to the following "This temporal lag between *R* and *GPP* meant that the forest floor on all plots had a greater tendency towards being a large net C source in the middle of the growing season" (Lines 260-262). We prefer this formulation because it is consistent with the general use of "forest floor" throughout the manuscript, and includes the contribution of not only understory and soil, but also tree roots.

Comment) P14Fig. 2. Use same scale/resolution in all figures (Figure 2a differs from the others), add notations for figures a, b, c and move axis legend of figure 2b to same side as figures 2a and 2c.
Response) Done.

Comment) Figure 3 text, P 14103L12 and L18, Hasselquist et al 2012 or 2010 or two different references? References list mention only 2012. Figure 3. Calculations of Ra and reasoning in text not clear.
Response) Thank you for pointing out this mistake. There is only one Hasselquist reference cited in the manuscript (Hasselquist et al. 2012), the 2010 citations have been corrected.

Reviewer 2 Comments

Comment) In addition, I think the correlation analysis is not correctly done. Correlations seem to be done over all 3 measurement periods. That is not correct if you want to compare biotic vs. abiotic explaining variables; there is large variation in fluxes over the measurements periods while the biotic factors tree proximity, shrub stem density and organic layer thickness do not change over time. Then it is not surprising that the biotic variables explain only little of the variance in fluxes. So correlations need to be done for each measurement period or for fluxes averaged over all measurement periods, so that n = 3 x 17 = 51 locations.
Response) These biotic factors may not change over time but they show very substantial spatial variation which is incorporated in the correlation analyses, so it is not clear that the correlations as we performed them will necessarily underestimate the role of these variables. Nevertheless, we agree that it would be interesting to partition spatial and temporal controls on the forest floor CO_2 fluxes we measure. So we have re-run the correlation analyses, as you suggest, for each time period to examine the effects of spatial variation in the abiotic and biotic variables measured on CO_2 fluxes, and revised Table 2 accordingly. To address temporal controls, we first controlled for spatial variation amongst points by dividing values for each location by the mean across all sampling periods for the same location. From this standardized data, we then calculated the slope of change in CO_2 fluxes, temperature, moisture and vegetation greenness for all locations across the three time periods, then ran correlation analyses on these slopes. We now include additional text in the methods detailing this approach (Lines 211-217), and have added the results of this new analysis to a revised version of Table 2. The results section has been updated according to these new correlations: as you suggest, the new analyses find a greater role for vegetation density in controlling spatial patterns of forest floor fluxes (Lines 267-269), though we could find no relation between temporal shifts in vegetation greenness and CO_2 fluxes (Lines 258-260, 296-300).

Comment) The experimental setup needs more explanation: How was the nitrogen added? Why were replicated nitrogen addition treatments not feasible? Is there information which shows that the three 15-ha plots were similar at the start?

Response) We now include additional text in the methods describing how nitrogen was added to the plots (Line 126). We now explicitly address the issue of the lack of true replication within our study in the methods section (Lines 128-131). Forest floor processes are strongly affected by the activity of the overlying trees, whose roots may extend up to 30 meters (Stone & Kalisz 1991) and possibly further for large horizontal tap roots. Therefore, to impose changes in forest nitrogen availability, we had to impose large scale fertilization treatments which we were not able to replicate for logistical and financial reasons. While we recognize the limitations of this approach, the values of such large-scale manipulations are widely recognized within the scientific community (Carpenter 1996, 1998; Sullivan 1997; Osmond et al. 2004), and results from un-replicated large-scale experiments (including the experiment in our study, Hasselquist et al. 2012) have made a large number of important contributions to the scientific literature on ecosystem ecology (e.g: see for example the literature emerging since 1955 from the un-replicated Hubbard brook watershed: http://www.hubbardbrook.org/). Even Steven Hurlbert, in his seminal paper on pseudoreplication (Hurlbert 1984), states "*The most common type of "controlled" experiment in field ecology involves a single "replicate" per treatment. This is neither surprising nor bad. Replication is often impossible or undesirable when very large-scale systems (whole lakes, watersheds, rivers, etc.) are studied. When gross effects of a treatment are anticipated, or when only a rough estimate of effect is required, or when the cost of replication is very great, experiments involving unreplicated treatments may also be the only or best option*". Hence, we believe that our experimental design, although unreplicated, provides useful insights into large-scale ecosystem responses to nitrogen availability which would have been impossible to capture in smaller scale, more easily replicated experiments. We encourage smaller-scale experimentation with forest floor responses to fertilization as an important complement to our approach, but caution that smaller-scale experimentation, whilst potentially more statistically rigorous, can lead to a range of misleading conclusions when results are scaled to the whole ecosystem (Tilman 1989, Carpenter 1996).

Nevertheless, in accordance with comments by yourself and reviewer 3, that the 17 measurements per plot are pseudoreplicates, we have decided to remove statistical analyses of plot-level differences because these analyses assumed that the 17 measurements were independent replicates. Instead, we present error bars derived from the 17 measurements to give an estimate of within-plot variability, as a guide to how robust the measured mean differences amongst the plots and seasonal measurements are.

We have measurements of tree production prior to fertilization, which we now present in the methods section (Lines 131-143) and Table 1. The results show that stem production on the fertilized plots was similar to the control prior to fertilization, but accelerated relative to the control after fertilization (Table 1). Given that stem biomass and production are important integrators of multiple forest processes (Gamfeldt et al. 2013), we interpret this as reasonable evidence that tree and understorey vegetation structure and composition were broadly similar across plots prior to fertilization.

Carpenter SR. 1996. Microcosm experiments have limited relevance for community and ecosystem ecology. Ecology 77: 677–680.

Carpenter, S.R. 1998. The need for large-scale experiments to assess and predict the response of ecosystems to perturbation. In: Successes, limitations and frontiers in ecosystem science (eds. Pace, M.L. & Groffman, P.M.). Springer-Verlag, New York, USA. pp. 287-312.

Gamfeldt, L. et al. 2013. Higher levels of multiple ecosystem services are found in forests with more tree species, Nature Communications, 4, 1340, doi:10.1038/ncomms2328, 2013.

Hurlbert SH. 1984. Pseudoreplication and the design of ecological field experiments. Ecological Monographs 54: 187–211.

Osmond, B. et al. 2004. Changing the way we think about global change research: scaling up in experimental ecosystem science. Global Change Biology 10:393-407.

Stone, E.L., & Kalisz, P.J. 1991. On the maximum extent of tree roots. Forest Ecology and Management 46: 59–102.

Sullivan, T.J. 1997. Ecosystem manipulation experimentation as a means of testing a biogeochemical model. Environmental Management 21:15-21.

Tilman, D. 1989. Ecological experimentation: strengths and conceptual problems. In: Long-Term Studies in Ecology (eds. Likens, G.E). Springer ,New York, USA, pp. 136-157

Comment) Ra, which is needed to discuss the carbon balance of understorey vegetation, as measured R includes heterotrophic R and tree belowground R, needs more explanation: How was Ra measured or calculated? Was it measured in another year?

Response) This section of the analysis utilizes published data on understorey vegetation respiration (R_{uv}) in Hasselquist et al. (2012) from the same plots and time periods as our study. This was already stated, but is now clarified further, in the introduction, methods, results sections, and in the Figure 3 legend (Lines 106-109, 219-221, 278-281, 603-605). We now take care to refer to the fluxes measured by Hasselquist et al. (2012) as R_{uv} (for understorey vegetation) throughout the manuscript rather than R_a (for autotrophic) which we acknowledge had been a source of confusion because tree roots are also a source of autotrophic respiration. We also now include a substantial extra section in the methods section summarizing the methodology of the Hasselquist et al. (2012) study and describing how their data were used for comparison with our data (Lines 219-233). To calculate understorey vegetation, Hasselquist et al. (2012) used plastic collars to trench areas of the understorey, thereby excluding contributions from tree roots. They then removed understorey vegetation from one collar but not another, and calculated understorey vegetation respiration as the difference in R between these two collars, thereby excluding the contributions of heterotrophs. They present understorey vegetation respiration estimates over 10 months after collar installation, so potential biases from disturbance should be minimized. This basic methodology has been successful applied in many published studies across a diverse range of ecosystems (Bond-Lamberty et al. 2004; Bowden et al. 1993; Haynes & Gower 1995; Subke et al. 2006).

In our analysis, we mistakenly took R values from the Hasselquist et al. (2012) analysis which represented contributions from both understorey vegetation, tree roots and extramatrical ectomycorrhizal hyphae. We have now re-calculated "plant carbon surplus potential" using estimated R contributions only from understorey vegetation and redrawn Figure 3 using this corrected data. The difference between the original and corrected versions are relatively minor (see figure at the end of this document) and do not alter any of our key conclusions.

Bond-Lamberty, B et al. 2004. Contribution of root respiration to soil surface CO(2) flux in a boreal black spruce chronosequence. Tree Physiology 24:1387-1395.

Bowden, R. D. et al. 1993. Contributions of aboveground litter, belowground litter, and root respiration to total soil respiration in a temperate mixed hardwood forest. Canadian Journal of Forest Research 23:1402-1407.

Haynes, B. E. & Gower, S. T. 1995. Belowground carbon allocation in unfertilized and fertilized red pine plantations in northern Wisconsin. Tree Physiology 15:317-325.

Subke J. –A et al. 2006. Trends and methodological impacts in soil CO_2 efflux partitioning: A metaanalytical review. Global Change Biology 12:921-943.

Comment) There is hardly any discussion of the results themselves: what is your explanation for the observed N treatment effects? Why is GPP in July so low? What is the role of the mosses?

Response) We did not measure moss abundance or CO_2 fluxes in this study, so we feel that discussion of the role of mosses on forest floor CO_2 fluxes at the site would be unwarranted. We have now included some additional text in the discussion to describe some reasons for the N treatment effects and the seasonal trend in GPP (Lines 296-301, 337-341). We are hesitant to discuss all of the results in the discussion, since this will take too much space and largely repeat the results section. Instead we view the discussion as an opportunity to select the most notable overall patterns and discuss them in more detail and in their broader context. Hence, we discuss at length what are, to us, the most interesting/important results of this study: **(1)** The generally low GPP, particularly compared to understorey vegetation respiration (Lines 293-332), **(2)** The variability of the N effects on CO_2 fluxes over the growing season, and (Lines 333-341), **(3)** The differences and similarities between our results and those of Hasselquist et al. (2012), and potential reasons for them (Lines 341-349), **(4)** The relation between CO_2 fluxes and some understorey species but not others (Lines 350-368).

Comment) p.14099, l.16-18 Isn't it possible to test for spatial autocorrelation?

Response) Yes, but it requires very extensive sampling (e.g.: Fortin, M et al. 1989; Philip Robertson, G. 1993) which would have been a substantial study in itself, and thus was beyond the scope of the current study. This sentence was a justification for treating the 17 measurements per plot as independent replicates, but based upon concerns about pseudoreplication by yourself and reviewer 3, we have removed plot-level statistical comparisons, and hence this sentence too.

Fortin, M et al. 1989. Spatial autocorrelation and sampling design in plant ecology. Vegetatio 83:209-222.

Philip Robertson, G. 1993. The spatial variability of soil resources following long-term disturbance. Oecologia 96:451-456.

Comment) Table 1. Please include letters to indicate significant treatment differences.

Response) In accordance with comments by yourself and reviewer 3, that the 17 measurements per plot are pseudoreplicates, we have decided to remove statistical analyses of plot-level differences because these analyses assumed that the 17 measurements were independent replicates. Instead, we present error bars derived from the 17 measurements to give an estimate to how robust the measured mean differences amongst the plots and seasonal measurements are.

Comment) Fig. 1. I think it is sufficient to mention the regression characteristics in the text, instead of this figure.

Response) We have removed these values from the figure and placed them in the legend (Lines 579-584).

Comment) Fig. 2. Why not include the overall treatment effect in the graphs? Or note in the legend that the overall treatment effect was not significant for all three fluxes.
Response) In accordance with comments by yourself and reviewer 3, that the 17 measurements per plot are pseudoreplicates, we have decided to remove statistical analyses of plot-level differences because these analyses assumed that the 17 measurements were independent replicates. Instead, we present error bars derived from the 17 measurements to give an estimate to how robust the measured mean differences amongst the plots and seasonal measurements are.

Reviewer 3 Comments

Comment) I have some problem with the experimental design of the study. There is ample replication of flux measurements within three areas of different N fertilisation (control andtwo N addition levels). This means that the main treatment (N addition) is not truly replicated, but rather that a large number (17) of sub-replicates are used for statistical analysis. I could not see any evidence that the plots were otherwise comparable prior to N addition was started, which is critical here.
Response) We now address the issue of the lack of true replication within our study in the methods section (Lines 128-131). Forest floor processes are strongly affected by the activity of the overlying trees, whose roots may extend up to 30 meters (Stone & Kalisz 1991) and possibly further for large horizontal tap roots. Therefore, to impose changes in forest nitrogen availability, we had to impose very large scale fertilization treatments which we were not able to replicate for logistical and financial reasons. While we recognize the limitations of this approach, the values of such large-scale manipulations are widely recognized within the scientific community (Carpenter 1996, 1998; Sullivan 1997; Osmond et al. 2004), and results from un-replicated large-scale experiments (including the experiment in our study, Hasselquist et al. 2012) have made a large number of important contributions to the scientific literature on ecosystem ecology (e.g: see for example the literature emerging since 1955 from the un-replicated Hubbard brook watershed: http://www.hubbardbrook.org/). Even Steven Hurlbert, in his seminal paper on pseudoreplication (Hurlbert 1984), states "*The most common type of "controlled" experiment in field ecology involves a single "replicate" per treatment. This is neither surprising nor bad. Replication is often impossible or undesirable when very large-scale systems (whole lakes, watersheds, rivers, etc.) are studied. When gross effects of a treatment are anticipated, or when only a rough estimate of effect is required, or when the cost of replication is very great, experiments involving unreplicated treatments may also be the only or best option*". Hence, we believe that our experimental design, although unreplicated, provides useful insights into large-scale ecosystem responses to nitrogen availability which would have been impossible to capture in smaller scale, more easily replicated experiments. We encourage smaller-scale experimentation with forest floor responses to fertilization as an important complement to our approach, but caution that smaller-scale experimentation, whilst potentially more statistically rigorous, can lead to a range of misleading conclusions when results are scaled to the whole ecosystem (Tilman 1989, Carpenter 1996).

Nevertheless, in accordance with comments by yourself and reviewer 2, that the 17 measurements per plot are pseudoreplicates, we have decided to remove statistical analyses of

plot-level differences because these analyses assumed that the 17 measurements were independent replicates. Instead, we present error bars derived from the 17 measurements to give an estimate of within-plot variability, as a guide to how robust the measured mean differences amongst the plots and seasonal measurements are.

We have measurements of tree production prior to fertilization, which we now present in the methods section (Lines 131-143) and Table 1. The results show that stem production on the fertilized plots was similar to the control prior to fertilization, but accelerated relative to the control after fertilization (Table 1). Given that stem biomass and production are important integrators of multiple forest processes (Gamfeldt et al. 2013), we interpret this as reasonable evidence that tree and understorey vegetation structure and composition were broadly similar across plots prior to fertilization.

Carpenter SR. 1996. *Microcosm experiments have limited relevance for community and ecosystem ecology. Ecology 77: 677–680.*

Carpenter, S.R. 1998. *The need for large-scale experiments to assess and predict the response of ecosystems to perturbation. In: Successes, limitations and frontiers in ecosystem science (eds. Pace, M.L. & Groffman, P.M.). Springer-Verlag, New York, USA. pp. 287-312.*

Gamfeldt, L. et al. 2013. *Higher levels of multiple ecosystem services are found in forests with more tree species, Nature Communications, 4, 1340, doi:10.1038/ncomms2328, 2013.*

Hurlbert SH. 1984. *Pseudoreplication and the design of ecological field experiments. Ecological Monographs 54: 187–211.*

Osmond, B. et al. 2004. *Changing the way we think about global change research: scaling up in experimental ecosystem science. Global Change Biology 10:393-407.*

Stone, E.L., & Kalisz, P.J. 1991. *On the maximum extent of tree roots. Forest Ecology and Management 46: 59–102.*

Sullivan, T.J. 1997. *Ecosystem manipulation experimentation as a means of testing a biogeochemical model. Environmental Management 21:15-21.*

Tilman, D. 1989. *Ecological experimentation: strengths and conceptual problems. In: Long-Term Studies in Ecology (eds. Likens, G.E). Springer ,New York, USA, pp. 136-157*

Comment) More significantly still, the approach fails to separate understorey gas exchange from tree root fluxes. N addition will have impacted the tree C allocation and hence belowground respiratory flux (see citations in the manuscript on this, mainly from the Hogberg group). By measuring net gas exchange and light response in these plots, a variable amount of C from tree roots is part of the measured response. I was surprised that throughout the manuscript all fluxes (assimilation and respiration) are interpreted as understorey fluxes only, when clearly the net C sink/source terms can only be obtained once the tree effect has been removed. As tree belowground allocation of C is a significant component of total CO_2 flux and also seasonably variable, I find it impossible to draw any conclusions regarding the understorey C balance from data shown in Fig. 2.

Response) We acknowledge that our use of the term "understorey" in the manuscript was misleading because we actually measure the contributions not only of vegetation, but also tree roots and soil. We did acknowledge this in the original manuscript (Page 14098, Lines 11-12), but nevertheless, to clarify this issue, we have replaced "understorey" with "forest floor" throughout the manuscript. This term gives a clearer description of what we actually measured, and has been used in a large number of published papers making similar measurements to our

study (e.g.: Drewitt et al. 2002; Gaumont-Guay. et al. 2014; Jassal. et al. 2005; Kelliher. 1999; Morén & Lindroth 2000; Swanson & Lawrence 2001). We now also take care to refer to the fluxes measured by Hasselquist et al. (2012) as R_{uv} (for understorey vegetation) rather than R_a (for autotrophic) which we acknowledge had been another source of confusion because, as you point out, tree roots are also a source of autotrophic respiration.

Quantifying the magnitude of, and environmental controls on, this integrated flux (incorporating ground vegetation, tree roots and soil) is of considerable scientific value because this information can be easily directly compared with other integrated fluxes (e.g.: whole forest net ecosystem exchange and respiration) and incorporated into ecosystem models with minimum error. This is a widely accepted approach, highlighted by the large number of published studies using these types of measurements (e.g.: Drewitt et al. 2002; Gaumont-Guay. et al. 2014; Jassal. et al. 2005; Kelliher. 1999; Morén & Lindroth 2000; Swanson & Lawrence 2001). Of course, partitioning the flux into its component parts is also important because it provides complementary, process-level information but this also usually introduces substantial uncertainties (e.g.: from trenching and/or soil disturbance). To then estimate integrated fluxes (e.g.: of forest floor respiration) from these partitioned data the individual errors associated with each of the component estimates would have to be summed. The final propagated error would likely be extremely large, compared with the direct measurement of the integrated flux that we perform in our study. In addition, while it is interesting to record the responses of individual soil components to environmental variation (e.g: in soil fertility, vegetation density or temperature) it is not straightforward to integrate these responses to understand the sensitivity of the whole system to the same variation because the components may interact together in difficult to predict ways. In conclusion, detailed partitioning of forest floor CO_2 fluxes into its components clearly has both strengths and weaknesses, but we strongly believe that our measurements of the integrated forest floor CO_2 flux bring valuable, complementary information worthy of publication.

Drewitt, G. B. et al. 2002. *Measuring forest floor CO_2 fluxes in a Douglas-fir forest. Agricultural and Forest Meteorology 110:299-317.*

Gaumont-Guay, D. et al. 2014. *Eight years of forest-floor CO_2 exchange in a boreal black spruce forest: Spatial integration and long-term temporal trends. Agricultural and Forest Meteorology 184: 25-35.*

Jassal, R. et al. 2005. *Relationship between soil CO_2 concentrations and forest-floor CO_2 effluxes. Agricultural and Forest Meteorology 130:176-192.*

Kelliher, F.M. 1999. *Carbon dioxide efflux density from the floor of a central Siberian pine forest. Agricultural and Forest Meteorology 94:217-232.*

Morén, A. & Lindroth, A. 2000. *CO_2 exchange at the floor of a boreal forest. Agricultural and Forest Meteorology 101:1–14*

Swanson, R. V. & Lawrence, B. 2001. *Environmental regulation of carbon dioxide exchange at the forest floor in a boreal black spruce ecosystem. Agricultural and Forest Meteorology 108:165-181.*

Comment) Interpreting flux differences as "plant carbon surplus" as in Fig. 3 is misleading. In the absence of any partitioning (why no trenching to exclude tree roots in plots???), I do not think that the conclusions are safe, and unfortunately conclude that a rejection of the manuscript is appropriate.

Response) This section of the analysis utilizes published data on understorey vegetation respiration in Hasselquist et al. (2012) from the same plots and time periods as our study. This was already stated, but is now clarified further, in the introduction, methods, results sections, and in the Figure 3 legend (Lines 106-109, 219-221, 278-281, 603-605). We now take care to refer to the fluxes measured by Hasselquist et al. (2012) as R_{uv} (for understorey vegetation) rather than R_a (for autotrophic) which we acknowledge was another source of confusion because tree roots are also a source of autotrophic respiration. We also now include a substantial extra section in the methods section summarizing the methodology of the Hasselquist et al. (2012) study and describing how their data were used for comparison with our data (Lines 219-233). To calculate understorey vegetation respiration, Hasselquist et al. (2012) did exactly as you suggest, they used plastic collars to trench areas of the understorey, thereby excluding tree roots. They then removed understorey vegetation from one collar but not another, and calculated understorey vegetation respiration as the difference in CO_2 efflux between these two collars. They present understorey vegetation respiration estimates over 10 months after collar installation, so potential biases from disturbance should be minimized. This basic methodology has been successful applied in many published studies across a diverse range of ecosystems (Bond-Lamberty et al. 2004; Bowden et al. 1993; Haynes & Gower 1995; Subke et al. 2006).

In our analysis, we mistakenly took R values from the Hasselquist et al. (2012) analysis which represented contributions from understorey vegetation, tree roots and extramatrical ectomycorrhizal hyhae. We have now re-calculated "plant carbon surplus potential" using estimated R contributions only from understorey vegetation and redrawn Figure 3 using this corrected data. The difference between the original and corrected versions are relatively minor (see figure below) and do not alter any of our key conclusions.

Bond-Lamberty, B et al. 2004. *Contribution of root respiration to soil surface CO(2) flux in a boreal black spruce chronosequence. Tree Physiology 24:1387-1395.*

Bowden, R. D. et al. 1993. *Contributions of aboveground litter, belowground litter, and root respiration to total soil respiration in a temperate mixed hardwood forest. Canadian Journal of Forest Research 23:1402-1407.*

Haynes, B. E. & Gower, S. T. 1995. *Belowground carbon allocation in unfertilized and fertilized red pine plantations in northern Wisconsin. Tree Physiology 15:317-325.*

Subke J. –A et al. 2006. *Trends and methodological impacts in soil CO2 efflux partitioning: A metaanalytical review. Global Change Biology 12:921-943.*

Month of measurement