Dear Referee one. We thank you for the constructive and detailed comments and suggestions. We agree with most of the comments and have made some further analysis to improve the study. We have made most of the required updates to the manuscript.

Major comments:

1. Using the mean GPP over no-outbreak years as the basis to estimate GPP reduction caused by outbreaks seems inadequate for two different reasons. First, differences in weather may affect both GPP and outbreak occurrence/severity. If the two variables are indeed correlated across years, the approach likely causes a bias. This should be checked; no need for fancy statistical tests, just compare, based on the GPP_lue model, the mean GPP and its standard deviation for outbreak vs. no-outbreak years over pixels that have never been defoliated. Second, GPP in pixels previously defoliated is unlikely equal to what it would have been if no outbreak had occurred. For example, canopy trees might have not fully recovered yet (hence underestimating the no-outbreak GPP) or total tree+understory productivity might increase for a few years due to the defoliation-caused growth release of the understory (hence overestimating the no-outbreak GPP). The authors should rather have estimated the no-outbreak GPP in pixels that have been defoliated up to X years before (with X to be defined; maybe 3-5 years?) based on the NDVI_DL values of neighbouring pixels that have never been defoliated. (More precisely: for each defoliated pixel, define a window large enough to include never-defoliated pixels, but small enough to have similar conditions. Then, compute the mean NDVI_DL difference between the defoliated pixel and the never defoliated pixels over all years prior to the (first) defoliation event in the defoliated pixel; let's say NDVI_DL was on average 10% higher in the defoliated pixel. Finally, for the X years after the defoliation event (excluding the defoliation year itself), estimate the annual no-outbreak GPP in the defoliated pixel with the GPP_lue model, but with a value of NDVI_DL 10% higher than the mean annual NDVI_DL value over the neighbouring never-defoliated pixels [instead of using the annual MODIS NDVI_DL value in the defoliated pixel].) Ideally, the authors should re-do their analyses using this new approach. At a minimum, the authors must use this new approach for >100 randomly-selected defoliated pixels and see to which extent it affects their results.

Response: We agree that these are import points. We first respond to the issue with recovery after an outbreak and thank you for the good suggestion about how to handle this limitation. We did explore the possibility to perform the suggested analysis but decided not to go ahead with the analysis. The main motivation is the difficulty to find pixels without any defoliation. Even though there are pixels that are not detected as defoliated during outbreak years we cannot know that these pixels are not influenced by lower defoliation levels. Slightly lower NDVI values may be due to meteorological conditions, but also due to minor defoliation by small larvae populations. Since it is not possible to distinguish between the two causes in remote sensing data such an analysis would increase uncertainties. Instead we modeled GPP based on PAR for the five years for which we have data from the EC-tower and compared to measured GPP. A comparison between measured GPP and PAR-modeled GPP suggests that the birch forests were slightly defoliated by growing larvae populations the two years prior to the outbreak in 2012. In 2007 and 2009 measured and PAR-modeled GPP agreed well. In 2007, measured GPP was slightly lower than PAR-modeled GPP and in 2009 measured GPP was slightly larger than PAR-modeled GPP. We have also observed lower NDVI values in the years prior to an outbreak in time-series of NDVI over the study area. One important note here is that the annual GPP values in Table 1 (p. 15 in manuscript) were not displayed in the correct order for the years 2009, 2010, 2011 and 2014. The correct numbers in Table 1 are given below and agree with the lower annual GPP values in 2010 and 2011:

	Years without insect outbreak					Outbreak
Year	**2007**	**2009**	**2010**	**2011**	**2014**	**2012**
GPP (g C m^{-2} yr^{-1})	451	531	373	401	448	180

For the outbreak year 2012 the difference between EC derived GPP and PAR-modeled GPP was 286 g C m^{-2} yr^{-1}, which is close to the decrease of 261 g C m^{-2} yr^{-1} estimated in the study. In addition, we ran the LUE model with meteorological data from the scientific research station in Abisko (ANS) for the year 2008 to fill the gap in the time-series with measured GPP and to study how well it agreed with the years 2007 and 2009. According to the LUE model the annual GPP at the EC tower was 440 g C m^{-2} yr^{-1} in 2008, which indicates that the GPP for undisturbed years of 441 g C m^{-2} yr^{-1} that we use is a reasonable value. We have added a discussion about these uncertainties and the challenge to find baseline conditions for GPP in areas with reoccurring insect outbreaks. In addition, a figure showing EC-derived and PAR-modeled GPP was added to the supplementary material. We have also added references that have found that the birch forests appear to reach pre-outbreak LAI and GPP 2-3 years after an outbreak (Hoogesteger & Karlsson 1992; Michal 2012).

As a response to the first part of the comment above we studied correlations between NDVI and meteorological data available from ANS, where we used the mean of the highest seasonal $NDVI_{DL}$ value derived from 200 MODIS pixels with birch forest. To minimize the influence of insect induced defoliation we excluded the outbreak years and years immediately prior to and after outbreaks. No linear correlations between PAR and GPP were found. There were, however, negative correlations between temperature and $NDVI_{DL}$, with the strongest correlation between $NDVI_{DL}$ and the mean temperature in May-June. The influence of temperature on $NDVI_{DL}$ was however, weak. Due to the low influence on NDVI and the large estimated uncertainty of the LUE model (30%) we did not adjust for these correlations in the analysis. We do, however, mention these results in the discussion but due to the limited amount of data we do not further elaborate on the results as that would be speculations.

There are studies related to insect outbreaks and climate but results are partly contradicting and only weak correlations between climate variables and outbreaks are found (se e.g. Young et al. 2014 and references within). Hence, we did not include this in the manuscript.

2. In the Discussion, the authors must at least explicitly acknowledge four major methodological weaknesses; when possible, explain the likely impact of each weakness (i.e., under- or overestimating defoliation-caused GPP losses) and propose a way to address the weakness. (1) fAPAR was based on measurements for the upper canopy only, so it is unclear to which extent the fAPAR vs. NDVI_DL relationship applies to the entire forest. This is particularly critical due to the (possible) growth release of the understory highlighted by the authors. (2) The fAPAR vs. NDVI_DL relationship was derived for undefoliated years (2010-2011) only, yet was also used in the GPP_lue,defoliated model. Why not developing a fAPAR vs. NDVI_DL for defoliated conditions (no defoliation event at the spectral tower over the entire study)? (3) The defoliation detection algorithm missed 26% of defoliated areas and misclassified 39% of undefoliated areas. (4) How representative was the EC tower footprint of the entire study area, both during outbreak and no-outbreak years? This is critical, as EC tower data provided the basis for all GPP estimates through the values of nepsilon_max, epsilon_max,def, and the GPP reduction factor.

Response: We agree that these weaknesses need to be discussed and have added them as limitations in the discussion: (1) This limitation is not easily handled since considerably more data are required to derive fAPAR and NDVI relationships depending on understory responses. We have, however, clarified the limitation in the discussion mentioning that a model including different relationships between fAPAR and NDVI depending on understory responses will be complex. (2) Unfortunately, the larvae were disrupting the PAR-sensors during the outbreak; hence, we have no reliable fAPAR data for defoliated conditions. We have added a section about this limitation in the discussion. (3) We have added a section about the accuracy of the defoliation detection method in the discussion: "The accuracy of the defoliation detection method also influences the results of the study. The method missed 26% of the defoliated MODIS pixels and misclassified 39% of the undisturbed pixels as defoliated in the evaluation data used by Olsson et al. (2016). This implies that the defoliated areas in 2004 and 2013 were slightly overestimated, while the defoliated area in 2012 was likely underestimated. However, the impact on the total numbers is likely small. It should also be considered that 20% of the forests in the study area were excluded since they are located in MODIS pixels with < 50% forests cover. The detection accuracy of the method may also influence the spatial distribution of the defoliation in the resulting maps, but even though there is an associated uncertainty at pixel level the broader outbreak patterns are likely reliable." (4) We agree that this could be an important limitations but according to Heliasz (2012) GPP is relatively stable over the study area. We have clarified this in the discussion: "There are also uncertainties in how well the EC tower footprint represents the study area. Heliasz (2012) utilized a permanent EC tower as reference and a mobile EC tower to study variability in carbon exchange in the birch forests around Abisko and concluded that there were only minor differences in GPP at seven sites during the peak growing season in 2008 and 2009. Hence, we consider the EC-tower footprint to be representative for the study area."

*3. I object to providing the 3-year *total* GPP reduction caused by defoliation, as this inadequately inflate numbers. Please provide the 3-year *mean* reduction instead throughout the text, making it clear the reduction is for outbreak years only (not the mean values over all years since 2000): Abstract; P16, L15 to P17, L1; P17, L7-8; P19, L5-7; P19, L15; P21, L18-20.*

Response: We agree that the total reduction in GPP may inflate the numbers. Hence, we have updated the manuscript accordingly except for in the discussion where we want to keep the comparison between the total decrease in GPP for the three outbreak years with the mean annual GPP for years without defoliation. We did, however, clarify that the total reduction was for the three years: "The total decrease in regional GPP, due to the three insect defoliation events studied here was estimated to be 44.6 ±13 Gg C, which is of the same magnitude as the average annual regional GPP of 41.1 ±12 Gg C yr^{-1} for single years with no disturbances."

*4. P3, L15-26. I have various issues with the text from "Since near-linear" to "(Liljedahl et al. 2011)". First, it should be in the Methods; I suggest merging at the beginning of Section 2.3. Second, units are not provided for the variables and the equation is not numbered; please number *all* equations and provide units for *all* variables throughout the text (even when unitless; e.g., NDVI). Third, the sentence starting on L20 is cumbersome; if kept in Section 2.3, I suggest re-writing along the following lines: "The light use efficiency coefficient varies between vegetation types and the influence of meteorological conditions is accounted for through reduction factors for temperature and vapour pressure deficit [...]". Fourth, the sentence starting on L23 is an overstatement because: 1) temperature is not* *always the main limiting factor in cold climates (Nemani et al. 2003; Beer et al. 2010); and 2) neither Bergh et al. (1998) nor Lagergren et al. (2005) really tested for the impact of factors other than temperature in their studies that covered only two sites in Sweden. The authors should acknowledge that they assumed accounting for temperature only was sufficient in their study region, an assumption supported (but not demonstrated) by Bergh et al. (1998) and Lagergren et al. (2005). Fifth, delete the sentence starting on L25: water stress is a major limiting factor in some boreal and other forests, not just for "ecosystems dominated by non-vascular plants".*

Response: We have moved most of the section to Methods as suggested and re-written or removed some parts of the text. What remains in the Introduction is: "Since near-linear relationships between satellite derived vegetation indices and the fraction absorbed PAR (fAPAR) have been established (e.g. Asrar et al. 1984; Sellers 1987; Goward & Huemmrich 1992; Myneni & Williams 1994; Olofsson and Eklundh 2007), it is possible to create a LUE model driven by remote sensing data. Such a LUE model could be applied for...". Consequently, Section 2.3 is updated according to the suggestions. The equation and the sentence "ecosystems dominated by non-vascular plants" are removed. We have also added to the methods (Section 2.3) that: "We assumed that accounting for temperature only is sufficient in our study region, which is supported by Bergh et al. (1998) and Lagergren et al. (2005)."

5. P6, L12. Why wasn't fAPAR_canopy also smoothed with TIMESAT before the regression?

Response: $fAPAR_{canopy}$ used in the regression is the mean of daily $fAPAR_{canopy}$ over eight day periods, coinciding with the MODIS eight day periods. Hence, the $fAPAR_{canopy}$ values in the regression are already smooth and we did not want to risk introducing further artefacts by applying more smoothing. However, we have clarified in section 2.2.2 that we are working with eight day mean values: "Average $fAPAR_{canopy}$ over eight day periods, coinciding with the MODIS eight day periods, were computed and an ordinary least squares (OLS) regression was performed…."

6. P7, Figure 3. The TIMESAT smoothing removed a second NDVI 'peak' each year. Please explain what was the origin of this (wintertime?) second annual peak and why removing it was OK.

Response: The second peak occurs during the winter when there is no vegetation in the area. We have not studied the origin of the second peak which could be due to e.g. snow or darkness (no light in the winter season). Hence, we do not discuss the second peak, but we have added a clarification to the figure caption explaining that removing the second peak is OK: "There is small peak in raw NDVI (left) appearing each year. This peak appears during the winter when there is no vegetation in the study area and is hence, removed from the smoothed data (right)." (Left and right will be replaced by a and b according to comment #27).

7. P12, L4-6. I do not understand why Method 1 should not also capture the effect of refoliation: the EC tower data should account for the post-refoliation increase in GPP, no? Unless no refoliation occurred within the EC tower footprint after the 2012 outbreak?

Response: To our knowledge there was no refoliation around the EC tower in 2012, hence, the reduction factor represents an outbreak year without refoliation.

8. P14, Figure 6. Many readers will likely expect defoliation to substantially decrease NDVI (due to a much lower leaf area index (LAI)) and leave LUE barely affected, so they will question the defoliation results (small NDVI decrease, large LUE decrease). It would thus be helpful to explain that small reductions in NDVI are associated with large reductions in LAI (e.g., Wulder et al. 1998), while LUE can *substantially decrease for lower LAI because more leaves operate in the light-saturated portion of the photosynthesis curve (e.g., Medlyn 1998). Also, please indicate the weeks during which defoliation occurred.*

Response: We agree with this reasoning and have added a section to the discussion: "It may seem unexpected that the difference in $NDVI_{DL}$ between undisturbed and defoliated years was comparably low in relation to the difference in light use efficiency. Mean $NDVI_{DL}$ for the peak of the growing season was 0.78 for the five years with EC-data. In 2012 the highest value for $NDVI_{DL}$ was 0.63. The difference in maximum light use efficiency was larger with an ε_{max} of 1.85 ±0.36 g C MJ^{-1} for years without disturbance, and an $\varepsilon_{max, def}$ of 0.98 ±0.25 g C MJ^{-1} during defoliation. It is however, well known that NDVI saturates for high LAI and that small changes in NDVI can be associated with large changes in LAI (e.g. Myneni et al. 2002). The light use efficiency on the other hand can decrease substantially with lower LAI since more leaves will operate in the light-saturated portion of the photosynthesis (e.g. Medlyn 1998)." Defoliation usually occurs just after budburst (second part of June) and refoliation in late July or early August.

9. P15, Figure 7 and P16, Figure 8. Give the equations for the regression lines on Figures 7 and 8. The line seems pretty close to 1:1 with zero intercept in Figure 7 (hence no bias for no-outbreak years), but the GPP_lue,defoliation model in Figure 8 seems to underestimate GPP over most of the range of actual values (i.e., up to _2 gC mˆ-2 dayˆ-1); this should be added to the list of weaknesses in the Discussion (see comment #2).

Response: We have added the equations for the regressions line in the figure captions. For non-outbreak years the equation is: GPP_{lue} = -0.12 × 1.01GPP. We also added the following to the text: "The intercept is -0.11 and the slope is 1.01 indicating that there is no bias in LUE modelled GPP for years without outbreaks." in Section 3.2, and "The figure, with an intercept of -0.54 and a slope of 1.25 indicates that the LUE model underestimates GPP for lower values." in Section 3.3.1. Furthermore, we will elaborate on the topic as a weakness in the discussion: Figure 8 suggests that the LUE model for defoliation underestimates GPP for values lower than about 1.5 g C m^{-2} day^{-1}. This is a potential limitation of the LUE model developed for years with defoliation and would require EC-data from more years with defoliation to study. However, for this specific year the underestimated values from the LUE model are mainly due to a cold spring that resulted in a large reduction factor (f_{8day}). During the main growing season LUE modelled and EC-derived GPP agree well.

*10. P18, L7-10. Unless I am mistaken, the authors do not correctly interpret these results.If Method 2 captured the effect of refoliation whereas Method 1 did not, then Table 3 results should have a higher absolute value for refoliated than non-refoliated pixels *regardless* of the actual sign. This is what was obtained, suggesting that Method 2 did better capture refoliation's effect; however, the differences between refoliated and non-refoliated pixels are always within the error margin, so this better performance is marginal. (The sign of Table 3 values (negative for 2004 and 2012, positive for 2013) is not directly related to the effect of refoliation, but to the bias between the two methods.)*

Response: It is true that we made a mistake for 2013 when there is little difference between the two methods. Thanks for noting this. We have updated the manuscript accordingly: "For all years the mean differences in GPP loss (g C m^{-2} yr^{-1}) between the methods were lower for pixels that recovered later in the growing season. These results suggest that Method 2 captured some of the refoliation, though the differences are small and within the error margin."

11. P19, L26 to P20, L3. I do not think these explanations for Table 3 results (i.e, the difference in GPP loss between Method 1 and Method 2) are appropriate. The refoliation argument does not hold because, as noted by the authors, both 2004 and 2013 had high refoliation yet had opposite signs for Method 1 minus Method 2. Furthermore, the lower GPP losses in 2013 for Method 2 were basically the same for non-refoliated and refoliated pixels (Table 3). The argument of "uncertainties in nepsilon_max,def" does not seem appropriate either, because the same value was used for all years (so why a sign change in 2013 for Method 1 minus Method 2?). The argument of higher NDVI due to understory growth would work if this growth was stronger in 2013 compared to 2004 and 2012; are there reasons to believe this was the case? Here is another hypothesis that could account for the much lower mean GPP reduction in 2013 under Method 2 compared to all other values (and hence account for the sign change in 2013): could it be that growing conditions were better in 2013, thereby leading to higher f_8day and/or PAR_8day used in GPP_lue,defoliated compared to the mean f_8day/PAR_8day no-outbreak values used in GPP_lue?

Response: We do agree that the discussion in this paragraph was a bit weak. We have removed the part about refoliation and uncertainties in $\varepsilon_{max, def}$. We studied meteorological data and compared the seasonal development in NDVI for the years 2004 and 2013. The seasonal trajectories of NDVI suggest that the growing season was shorter and that refoliation started earlier in 2013, which is one possible explanation for the smaller decrease in annual GPP for Method 2. We have added this to the discussion: "For the years 2004 and 2012, the two methods resulted in similar estimates of the GPP loss with slightly larger decrease in GPP for Method 2. In 2013, the difference between the methods was larger with the highest decrease in annual GPP for method 1. One possible explanation for the smaller decrease in annual GPP according to Method 2 for the year 2013 is that the growing season seems to have been shorter and that refoliation started earlier and was stronger in 2013 compared to 2004; this is indicated by the seasonal developments of NDVI for the pixels around the EC-tower."

Minor comments:

We agree with most of the minor comments. To keep this response short we only include the minor comments that we want to give any specific response to. For all minor comments that are not listed below the manuscript has been updated accordingly.

12. Throughout the text: when possible replace the vague "carbon uptake" expression by the more accurate term applicable (GPP, NEE, etc.). Therefore, the title should be: "Mapping the reduction in gross primary productivity due to insect defoliation in subarctic birch forests". In the Discussion, the authors should stress that their results are not for the net carbon balance, but for GPP only; based on Heliasz et al. (2011), would it be possible to speculate whether, on a percentage basis, NEP losses should be higher or lower than GPP losses?

Response: We have changed to either GPP or NEE and accordingly also changed the title as suggested and we have stressed that the results are for GPP only. During the outbreak in 2012 the decrease in R_{eco} was much larger than decrease in GPP during the growing season around the EC tower.

15. P2, L1. Reference(s) should support the "a warmer climate is likely to increase forest productivity" part of the sentence. Here are some suggestions: Pastor and Post (1988), Nemani et al. (2003), or Boisvenue and Running (2006).

Response: We have added "(e.g. Nemani et al. 2003; Boisvenue & Running 2006)" as suggested.

19. P3, L27-28. Bright et al. (2013) already used remote sensing (Landsat to identify trees killed by the mountain pine beetle) and a LUE model (MODIS GPP results, which *are based on LUE) to quantify the impact of an insect outbreak on carbon uptake. To my knowledge, the authors can still claim being the first ones to do it for defoliators.*

Response: We have updated the manuscript accordingly and added Bright et al. (2013) as a reference.

23. P5, L18 to P6, L1. The purpose of "quality classes based on QA data" is not clear: how where these classes used in the analysis?

Response: We agree that mentioning the quality classes was confusing, hence, we have changed the text: "Double logistic functions were used to smooth the raw NDVI data and QA data from both MOD09Q1 and the more comprehensive QA-flags in MOD09A1 were utilized to estimate the quality of the NDVI observations."

25. P6, L8. Is the "pure canopy absorbed PAR" different from fAPAR_canopy? If yes, quickly explain what is the difference and why it matters. If not, replace the expression by "fAPAR_canopy".

Response: Pure canopy absorbed PAR is the same as $fAPAR_{canopy}$. We have updated the manuscript accordingly.

27. P7. I suggest combining Figure 2 with Figure 1. For all multi-panel Figures, please add letters to each panel and refer to the appropriate panel in the text (e.g., Fig. 5a).

Response: We do agree that it could be a good idea to merge the figures so that the orthophoto over the area around the EC-tower is shown together with the location of the tower. However, we decided to keep them as separate figures with the main motivation that we want to keep the size of Figure 2 large to make the photo easy to interpret. Letter have been added to multi-panel figures.

32. P10, L3. T_mean8 was never negative for such a northern site at the end of September? If negative values occurred, explain how they were dealt with (Equation (8) suggests a negative value for f_8day, which would make no sense).

Response: There were no negative values for T_{mean8} in the period studied (coldest value was 3°C). There were eight day periods with temperatures below zero, but no eight day period with a mean value < 0°C.

36. P11, L15-16. The text should be expanded, because at this point the reader is still unaware that an outbreak occurred in 2012 within the EC tower footprint: please state this explicitly here and add what was the percentage of defoliation within the tower footprint.

Response: We have clarified according to the comment: "(1) a method based on a reduction factor derived from the EC data from 2012 when the birch forest in the footprint of the tower was severely defoliated, and no refoliation occurred. This reduction factor was applied to all pixels in the study area"

48. P21, L1-2. For (partially) contrasting modeling results about the effect of insect outbreaks on the carbon balance, see Seidl et al. (2008), Albani et al. (2010), and Landry et al. (2016) (the last one is not the same reference as already cited in the manuscript).

Response: Thanks for these interesting references. We did not include them though as they mainly discuss longer term impacts.

50. P21, L13-14. Delete the sentence: there is no firm basis for such an extrapolation, and the number would then likely end up being cited by future studies...

Response: Since these insect infestations occur frequently across very large areas in the region we think that it is relevant to include some information about the potential effect on the carbon uptake. However, we agree that it is an uncertain estimation and reformulated the text accordingly: “Assuming that the conditions were similar over northern Fennoscandia, the insect defoliation over these vast areas would result in a potential total regional GPP loss for the time period of the magnitude 2–3 Tg C.”

References:

Heliasz (2012). PhD thesis, Dep. of Geography and Ecosystem Science, Lund Univ., Lund, Sweden.

Hoogesteger and Karlsson (1992). Functional Ecology, 6, 317-323, 10.2307/2389523.

Young et al. (2014). Arctic, Antarctic, and Alpine Research, 46, 659-668, 10.1657/1938-4246-46.3.659.