Dear Editors and Referees,

Thank you for your constructive and valuable comments! From the multitude of suggestions, we draw the general impression that the topic and the design of our study are timely and interesting. In the revised version we deleted the term “extreme” completely from the manuscript since most of the referees couldn’t agree on it. However, the altered winter conditions in our study are clearly more extreme in terms of temperature variability than the ambient reference conditions. We are convinced that taking increased variability into account when dealing with the importance of extremes is a very important consideration. Further we changed parts of the introduction, methods and discussion to clarify our rationales and the measured parameters. We now clarified that the bait method and PEEAs are proxies for biotic decomposition. We further state more clearly that a dilution effect of ^{15}N incorporation can be excluded. Some of the figures were also improved as suggested (Fig. 1 & 4).

Best regards

Jan Schürings and co-authors

Anonymous Referee #1 Received and published: 10 June 2014

1. General comments: The manuscript describes a study on the impact of warming pulses during winter on N cycling in middle Europe. The results are interesting but the introduction and discussion are very superficial and do not indicate how this study relates to the existing literature and how it extends/broadens the current knowledge on this subject. In addition, considering that the measurements included various variables that could have affect N cycling (soil biotic activity, plant species richness, biomass, diversity, plant identity) no attempt was made to compare the relative role of each of these factors or the potential interaction between them. Finally, the induced temperature regimes suggest that the warming was a more likely factor behind the N responses as there were hardly any differences in freezing temperatures between treatments. The intensity of the warming pulses and their realism, are however not discussed in the manuscript.
Response: We believe that just showing the importance of increased soil temperature fluctuations on several response parameters linked to the N-cycle is novel and important, because studies so far have mainly linked observed changes in the N-Cycle over winter to increased number of freeze thaw cycles and/or changes in minimum temperature. We were lucky enough that the number of freeze-thaw cycles and minimum temperature did not change, so that our data clearly suggests that variability itself due to warming also needs to be considered. From this and other comments below we saw that we had to improve the selling of this message! With regards to the relative importance of different variables, we agree that this will make an interesting contribution to the discussion. We did not do this so far because this is not strictly testable in this experimental setup, yet we added some observational discussion here. With regards to testing the interactive effects of our drivers (site, plant community, warming pulses, treatment) we have formally tested for interactions and displayed all results.

2. Specific comments Lines 11-14: line 11 suggests that the current knowledge on N uptake is fully addressed. However the two following lines indicate that this is far from clear. Please rephrase these lines. Further there is no explanation on the mechanisms of winter N uptake. Which processes are involved; what is their activity during winter compared to summer and how are these processes affected by temperature changes?

Response: In line 11 we exactly wrote that effectiveness of N uptake from plants during winter is not fully clear yet. Further we wrote that some studies found comparable N uptake rates during summer and winter but other studies showed reduced N uptake during winter. With higher temperatures plants can deacclimate and start getting active during winter (Kalberer et al. 2006: Deacclimation and reacclimation of cold-hardy plants: Current understanding and emerging concepts; Plant Science 171; 3–16).

3. Line 16: in the previous paragraph the focus is on soil temperatures why the shift to air temperature?
Response: Increasing winter air temperature is the generally accepted scenario with ongoing global warming. Yet it is also the reason behind decreased snow cover which then leads to more variable winter soil temperatures. So, we wanted to make clear that not decreasing precipitation (i.e. snowfall) is the reason behind decreasing snow cover but increasing winter air temperatures. We clarified this message which now states: "Due to increased winter air temperatures, snow cover will decrease in many regions of the temperate zone (Christensen et al., 2007; Kreyling and Henry, 2011). However, air frost events will still occur with unchanged magnitude and duration as nowadays in many temperate regions (Kodra et al., 2011), and with less insulating snow cover, winter soil temperatures can become more variable, particularly in upland and cold temperate regions (Henry, 2008; Brown and DeGaetano, 2011)."

4. Lnes 16-18 why is there a contradiction here 'nevertheless' ?
Response: See point 3

5. Line 18 as frequently as what? The whole paragraph starting at line 16 is incoherent and lacks logical steps between statements and conclusions.
Response: We reworded the paragraph. See point 3.

6. Line 5 (7800) please indicate as to how/in what way grasses are more responsive than dwarf shrubs and how does this reflect on the contrasting responses observed in the two cited studies? Thereafter, these differences in responsiveness need to be linked to N cycling which is the main theme of this paper. Is the positive/negative responsiveness of species/groups directly linked to the plants ability to take up N during winter?
Response: In the cited studies generally grasses (especially *H. lanatus*) showed a response to pulsed winter warming in form of decreased or increased productivity whereas dwarf shrubs showed no, weak or late responses. N cycling wasn't a focus of these studies but changes in productivity indirectly can affect N cycling. We now state: "Different plant species further show variability in their reactions to increased winter temperature variability in the temperate zone, with grasses appearing more responsive than dwarf shrubs regarding their productivity (Kreyling et al., 2010; Schuerings et al., 2014). However, this increased responsiveness in productivity can either be beneficial (Kreyling et al., 2008), or detrimental (Schuerings et al., 2014), probably depending on whether the minimum temperatures experienced after warm phases induce frost damage. Altered plant productivity can therefore indirectly affect N cycling."

7. Lines 11-12 why would increased N affect the increased risk of frost damage? One important article that needs to be cited here: Macgillivray CW, Grime JP, Band SR, Booth RE, Campbell B, Hendry GAF, et al. Testing predictions of the resistance and resilience of vegetation subjected to extreme events Funct Ecol. 1995;9(4):640-9.
Response: We have included this reference and further information (see ll. 83-85).

8. Line 15 where does the extracellular enzyme activity suddenly come from? This should be properly introduced in the previous introduction paragraphs. I don't understand the logic behind hypothesis 2. If a site is located in a region with typically colder winter temperatures why would you than expect it to respond stronger to changes in winter temperature than a site that has milder winter temperatures? I would expect species to adapt to the severity of the winter climate.
Response: In the introduction we tried to focus on general effects of winter warming/pulsed winter warming on N cycling. Therefore, specific response variables aren't introduced. We then use a set of response parameters which cover several aspects which influence the N-cycle. Extracellular enzyme activity is one of them. In line with reviewer 5, we will add short explanations on why and how the single response parameters are useful for the interpretation of the experiment. We did this in the introduction and also in the method section. The rationale for hypothesis 2 is as follows: The upland site is more snow rich and therefore less variable in winter soil temperatures. Loss of the snow cover by warming pulses should therefore lead to more variable soil temperatures at the upland site than at the naturally snow-poor lowland site. We improved the clarity of this rationale.

9. Lines 23-24 where do the grassland and heathland communities suddenly come from? These should be introduced in the previous introduction paragraphs.
Response: On p. 4, ll. 4-12, we introduce results of pulsed winter warming experiments with grassland and heath species. We clarified these results and linked them to the community compositions used later on (Schuerings et al., 2014).

10. Methods: Line 8 are you sure that the botanical gardens of Bayreuth are only 11° N of the equator?
Response: We corrected this mistake.

11. Line 12 how were the heating cables placed in relation to the soil? Did the heating cables in any way interfere with the growth of the plants? Were dummy heating cables also added to the ambient plots (similar to the lamps)? Why were the heating cables inserted and was there also a treatment with lamps but without soil heating cables? Were the measured soil temperature increases (nearly reaching 10 C) in coherence with observed/expected changes in soil climate?
Response: The heating cables lay on the soil surface with a distance of around 20cm to each other. Since the cables were placed on bare soil next to the plants no interference with plant growth is expected. We are here interested in the effects of warming pulses which take away the snow. Pre-trials showed that IR-lamps were insufficient for this purpose, especially when there was wind. There was no treatment with lamps only, because we were not interested in the difference of the warming source. Phenomenological, we argue that natural thaw events over winter which take away the snow cover are often linked with rainfall. The rain provides the heat for quick snow melt within the snow layer. The warming cables therefore acting within the snow layer should therefore act similarly to rain water.

12. Line 15 how many pulses were given and why that number?
Response: Six warming pulses were applied which is in the natural range of warming pulses possible in this region. The number was dependent of ambient weather conditions during the manipulation period since we needed 3 to 4 days of continuous frost at both experimental sites. We only applied warming pulses when soil was frozen and local weather forecast predicted at least 48h of continuous frost as otherwise we expected no real change in temperature variability. We now state in the corresponding sentence that we administered six warming pulses.

13. Line 21 why calculate an abstract mathematical value (CV) which is much harder to comprehend, for temperature when you have the actual temperature data? Simply state the temperature of one treatment to that of the other or calculate a relevant measure for temperature variability i.e., freeze-thaw cycling, occurrence of deeper frost etc.
Response: In the beginning of the result section (p.10 l. 19 - p. 11 l. 3) we give mean soil temperatures, minimum soil temperatures and number of FTC. However, our aim was to increase soil temperature variability, which cannot be expressed by these numbers. Even increased numbers of freeze-thaw cycles would not express increased variability per se, as the same variability would also be possible without crossing zero degree at all. Therefore, the CV is the appropriate index for our study showing that treatment plots had a higher variance in soil temperatures than control plots.

14. Line 23 so there was a snow camera for each of the mesocosm blocks?
Response: There was one snow camera for each site which took pictures of one treatment block and one ambient block. But since blocks were standing on a free field, no differences in snow height are expected. Visual inspections during all visits to the sites showed no difference between the camera-blocks and the respective treatment blocks.

15. Response parameters: Why was the bait lamina method used here? It has previously been shown that his method only works reasonably well for earthworms; Gestel CAM, Kruidenier M, Berg MP. Suitability of wheat straw decomposition, cotton strip degradation and bait-lamina feeding tests to determine soil invertebrate activity. Biology and Fertility of Soils. 2003;37(2):115-23.
Response: Based also on reviewer 4's comments, we will add a rationale for each response parameter. Concerning the bait-lamina test, it is a commonly used method (eg. Kratz 1998; Geissen and Brümmer 1999; Filzek et al. 2004) which can be realized with limited space and disturbance (two very important points in mesocosms). We included details which guilds feed on the baits in the methods section.

16. 'The stopping buffer also raises the pH of the solution, because fluorescence is strongest in the alkaline pH range (pH> 9).' Surely the stopping buffer does not raise the pH because fluorescence is higher at pH over 9. A logical step/explanation is missing here.
Response: We shortened the description of this method. It now states: "Three soil samples (2 cm diameter, 10 cm depth) per mesocosm were collected and mixed for assays of potential extracellular enzyme activity (PEEA) in soil on 21 February 2011. Soil samples were stored in airtight plastic zip-bags at 4°C and were analysed within 3 days. PEEA assays were carried out with Methylumbelliferone substrates (MUF) (Pritsch et al., 2004; Pritsch et al., 2005). The following PEEAs were measured: MU-β-D-glucopyranoside (MU-G), for β-glucosidase, MU-β-cellobioside (MU-C) for cellobiohydrolase, MU-β-D-xylopyranoside (MU-X) for xylosidase, MU-phosphate (MU-P) for acid phosphatase. Substrates and calibration saturation and incubation times were determined in pre-experiments (data not shown) as follows: MU-G and MU-X each 500 µM incubating for 60 min, MU-C 500 µM incubating for 120 min, MU-P 800 µM incubating for 40 min. Fluorescence was detected at an excitation wavelength of 360 nm and an emission wavelength of 450 nm with a Gemini EM Fluorescence Microplate Reader from Molecular Device, California."

17. Why were the enzyme essays done at 21° C? This is not a typical soil temperature for central Germany and certainly not typical for winter conditions. So how representative is this analysis for the actual activity in the field?

Response: This is not the actual extracellular activity, but the potential extracellular enzyme activity (PEEA). This method was done after a standardized common method to compare results to other experiments.

18. Line 20 here it is stated that the soil temperature became more extreme. The term ‘extreme’ needs a better definition. Further, on line 23 it is stated that the minimum temperature was nearly identical (differences of 0.2 C) between treatments this does not suggest a more ‘extreme’ soil temperature environment. There were no changes in freeze-thaw cycles and the ‘much colder temperatures’ were only 2 C lower than the control plots in the colder site.
Response: The temperature stated in line 23 is only the difference in minimum temperatures at the lowland site. For the upland site the difference is -2.1 C in minimum temperature. The lowland site showed a "strong" increase in mean temperature (+1.7 C) but only a small decrease in minimum temperatures (-0.2 C) whereas the upland site showed a small increase in mean temperatures (+0.2 C) but a "strong" decrease in minimum temperatures (-2.1 C). We used the term “more extreme” here to express that soil temperature variability was increased. We agree that this lacks any information on the absolute extremeness or rarity of the conditions. In order to avoid misinterpretation we now use the term “more variable soil temperatures” in the revision.

19. Discussion line 19: see earlier comments on the severity of the treatments. The treatments need to be viewed in relation to the natural variation that exists across years. Is -2 C the coldest ever measured at the ‘cold’ site?
Response: Our treatment was in the range of natural variation and colder soil temperature extremes have been measured at the site. Our point is that more variable soil temperatures and winter warm spells will become more frequent in the future and this will influence N cycling. We argue that absolute minimum temperature is not always the most important number. A frost of -5 C, for instance, kills organs of temperate plants during the growing season but has no effect in hardened winter states. We propose that hardening, and in particular quick dehardening after warm spells over winter can considerably change cold tolerance of organisms. In fact, we just got data from a new experiment which clearly demonstrates that winter warm spells can activate photosynthesis in plants, making them vulnerable to subsequent frost which would not have been damaging without activation. Simple consideration of minimum temperatures is clearly insufficient. For the interpretation of the results, however, we agree that it will be interesting to compare our treatments to long-term mean conditions at each site. Yet future conditions are unknown as climate change scenarios lack the necessary detail both in temporal resolution and in display of realistic minimum and maximum temperatures. The realism of the experiment with regard to future conditions can therefore not be evaluated. We can state, however: “While changed FTCs, higher mean soil temperatures and freezing intensity are important drivers of N cycling, our results show that with only increasing soil temperature variability N availability and soil biotic activity are increased.”

20. Lines 23-25 true statement but how does this reflect to your findings? The results indicate that there were no differences in FTC between treatments, suggesting that for your study sites FTC were not relevant. How does this relate to the results/conclusion of the cited articles? Could the strong responses measured in this study have anything to do with the soil temperatures reaching nearly 10 C?
Response: This is exactly the message: N-cycling over winter can be affected without changing FTCs. Yet it is not as simple as saying that the warm peaks alone have caused the observed effects, as the plant data demonstrated that the warm pulse treatment was

detrimental rather than beneficial. We emphasized this aspect more strongly, compare also to point 1.

21. Page 7809line 5 you already mentioned the role of FTC on N cycling. Lines 5-10 this discussion needs to be broadened; the results indicate no differences between treatments in FTC but increased cycling of N. Therefore, this suggests that FTC did not affect N cycling in this study, this needs to be reflected to the findings of others. Further, how does the -4 C stand out from longer term records? Why would the occurrence of a decreased minimum (which may only have occurred once; this is not made clear) by 2 C affect N cycling so much? What actually stand out from figure 1 is the much higher soil temperatures that were induced during winter and not so much the colder (freezing) temperatures following these events. This needs to be addressed in the discussion.
Response: We generally agree (see also point 1, 19 and 20). However, we cannot conclude that FTC are unimportant, we rather show that considerable effects happen even when FTC are unaffected. We inserted: "While changed FTCs, higher mean soil temperatures and freezing intensity are important drivers of N cycling, our results show that with only increasing soil temperature variability N availability and soil biotic activity are increased."

22. Line 15 how does the local climate affect N cycling?
Response: We cannot answer this question with our data.

23. PRS probes have a passive exchange of anions and cations with their environment meaning that at all times the PRS is taking up and releasing ions. The extracted NH4 and NO3 from the PRS probes therefore reflect the availability of these anions and cations right before sampling and do not incorporate what has happened during winter. (PRS probes do not capture and hold onto NH4 and NO3 indefinitely).
Response: We agree that this system cannot capture the temporal dynamics of N availability. Using it for short periods, however, provides an integration over this period of time rather than the snap-shot provided by soil sampling and extraction. In line with reviewer 4, we expanded the explanation of meaning and limitations of the response parameters in the methods section. Here, we used a test system from western AG innovations Inc. which has been used in many studies by now. Their recommendation was that at least 12 weeks burial time is okay during winter. The maximum ion capacity for nitrate is given with 2088 µg 10 cm-2. Our values are far lower, indicating that the system was not saturated. At the end of this paragraph in the method section we added: "The maximum ion capacity for nitrate is given with 2088 µg 10 cm^{-2}. The values in our study are far lower, showing that the system was not saturated. Since NO_3/NH_4 uptake by resin sticks is not a linear process, values are given for the whole burial period."

24. Lines 20- end of paragraph, this is mainly a summary of the results again without clear explanations of how and why differences were observed and there is no citation to any other literature in this paragraph.
Response: We agree partly. Summing up the results and starting an "internal" discussion still appears useful with regard to the complexity of our experiment.

25. Line 25 (7810) I have not read about any biomass measures in the results or methods section: : : This data needs to be incorporated in the results section or the text should be removed from the discussion on this subject.
Response: Meanwhile our manuscript on plant biomass of this experiment has been published and we specifically refer to this. See Schuerings et al. (2014) Winter warming pulses

differently affect plant performance in temperate heathland and grassland communities; Ecol Res; DOI 10.1007/s11284-014-1174-x.

26. What is the difference between chronic and continuous warming?
Response: There is no difference. We now only use the term chronic warming in the manuscript.

27. Conclusions: It is true that this study showed N cycling to respond to the variable temperatures induced by the treatments but I have some doubts on whether this is in response to the minimal changes in freezing temperatures; instead the much higher soil temperatures may have played a stronger role here. The difference in 15N uptake between sites does not follow logically from increased frost damage. There was no apparent larger frost apart from a one-off lower freezing temperature of just 2 C which I would expect to be part of the climatic variation. In the conclusion the authors also allude to frost damage but there was not data on shoot biomass of damage measured presented in the results; this conclusion is therefore not justified.
Response: This point reflects on several others above, please refer to our responses there. In short, we agree that the absolute minimum temperatures are not overly important. Yet, as stated already repeatedly, the temporal variability is crucial: Since plants can start dehardening after hours of elevated temperatures (Kalberer et al., 2006) a subsequent frost event can damage the plants. This means that neither the minimum nor the maximum temperatures alone are relevant, but only their temporal succession. A decrease in minimum temperatures from -2.6 C to -4.7 C may crossed a threshold value for plant frost damage of some species (3 out of 4). In the resubmitted version, the included proposed changes to the discussion are also be mirrored by a clarified conclusion.

28. Figure 1 considering that heating lamps and cables were present how come the temperature increases were so variable? In addition, some of the soil temperature treatments reached nearly 10 C this is somewhat worrying considering that all this heat had to come from above and may have seriously overheated the aboveground plant parts.
Response: Soil temperature reaching 10 C for a short period during winter is not out of the natural range for our region. The effects of the heating lamps depended on wind speed, fog, etc. That’s why temperature increases were variable. The air temperatures displayed in 1a are taken at plant height. We do not see an unrealistic “overheating” here.

29. Figure 2 please indicate which bars significantly differ from each other in the figures.
Response: We think that effect sizes matter much more than the mathematically correct but ecologically irrelevant question if two groups differ by a minimal amount larger than zero (which is strongly affected by the ecologically irrelevant number of replication or sample size). We are convinced that the effect sizes together with the information that the corresponding ANOVA yielded significant differences among groups is more clearly presented without indication of post-hoc test results. Consequently, we do not show any post-hoc tests in the manuscript.

30. Figure 4 bc suggest reducing the scale of the y-axis to improve readability.
Response: We changed this.

Anonymous Referee #2 Received and published: 21 June 2014

31. General comments: The manuscript entitled “Recurrent winter warming pulses enhance nitrogen cycling and soil biotic activity in temperate heathland and grassland mesocosms” by Jentsch et al. reports the interesting results about the context (i.e., vegetation type, habitat abiotic environment) dependency of the effect of extreme warming pulse on plant-soil systems. Overall, the manuscript is well-written. On the other hand, with the plenty of data that the authors have, to avoid the speculative discussion, it can be once choice to clarify the mechanism through which, for instance, the available N increase, by conducting the linear mixed model by setting several factors as explanatory variable (e.g. enzyme activity, plant N uptake: : :) and doing N availability as response variable. Furthermore, the number of zero-crossing in soil temperature is helpful to have the implication about the effects of soil FTC on the tested parameters.
Response: Thank you for the positive overall evaluation! We generally like the suggested addition to the analyses. However, we see several technical issues which prevented the suggested analysis, mainly the missing independence among the explanatory variables. We tested several options and decided to stick to our analysis.

32. Specific comments: P7799-Line 7: Why “However”? And I think “soil freeze-thaw cycles” would be proper here because, for example, Joseph and Henry (2008) study not the effect of thawing solely but that of freeze-thaw.
Response: See next point.

33. P7799 Line 7-9: This part can be moved to the second paragraph where soil FTC is described. But please be sure that line 25-26 is overlapping with this sentence.
Response: We incorporated the references from the sentence in the suggested lines.

34. P7800 Line 7: Please add short explanation about why the grasses are more responsive than shrub.
Response: In the cited studies generally grasses (especially H. lanatus) showed a response to pulsed winter warming whereas in form of decreased or increased productivity whereas dwarf shrubs showed no response. The reason for this, however, is far from clear but might be linked to the generally faster life cycle of the grasses. We added a half sentence on this.

35. P7800 Line 13: Add “of” just after “the effects”ãËŸA0C
Response: This mistake is corrected.

36. P7800 Line 11-12: This can be deleted because this information is not necessary to draw the hypothesis.
Response: In line with Referee #1 we included a reference with a possible explanation (see point 7).

37. P7801 Line 10: N could be E.
Response: We corrected this mistake.

38. P7809 Line 11: How about the possibility that the increase of N availability is caused by the reduction of N incorporation by pulse treatment? It is nice to move a part of P7810 Line 4-19 to this part.
Response: Indeed, this is what we want to express, so we added: “Since we found lowered N incorporation into plants (see discussion further down) and lower biomass (Schuerings et al., 2014) at the cold site, this could have lowered N immobilization by plants. Freezing intensity is therefore another important determinant of N cycling responses, possibly damaging dehardend plants.”

39. P7809 Line 21-22: I could not catch what the authors wanted to mention with this sentence. Please re-word.
Response: We reformulated the sentence. It now states: “Since there were no roots in the bare ground plots competing with the PRS^{TM}-probes for N, this result is not surprising.”

40. P7810 Line 10: Please insert “directly” between “damage” and “plants”.
Response: Thank you for the suggestion.

41. P7811 Line 2: It is nice to mention the importance of functional trait-base study to clarify the determinant of species specific response. Please read Cornelissen and Makoto (2014) Winter climate change, plant traits and nutrient and carbon cycling in cold biomes Ecological Research (DOI 10.1007/s11284-013-1106-1).
Response: We included this reference.

42. P7811: Line 15: Here Makoto et al. (2014) Winter climate change in plant–soil systems: summary of recent findings and future perspectives Ecological Research DOI: 10.1007/s11284-013-1115-0 pr
Response: We included this reference.

Anonymous Referee #3 Received and published: 23 June 2014

43. The article, “Recurrent winter warming pulses enhance nitrogen cycling and soil biotic activity in temperate heathland and grassland mesocosms,” attempts to experimentally test the responses of soil biotic activity, N cycling, and plant N uptake to repeated winter warming pulses in temperate heathland and grassland mesocosms situated in both a colder and warmer site. Although I think the premise of the study is of interest, there were a number of issues that make the logic of the paper fairly difficult to follow and jumps between data and interpretation that I feel ought to be more conservatively structured. In general, I found the logic of the introduction and discussion very difficult to follow. Many measurements were described, but few connections among these measurements were explained, which would have helped to give insight into how N cycling might change with warming and concomitant freeze thaw pulses.
Further, I have several major concerns stemming from the methodology of the experiment and subsequent interpretation: The experiment was set up as a mesocosm study, with soil substrate consisting of homogenized loamy sand from a nearby quarry. While I understand the advantages gained by mesocosm experiments, I am concerned about broad-scale interpretations of changes in soil N pools, plant N uptake, and soil biotic activity from this set up.
Response: Thank you for the overall positive evaluation of our study focus. Of course, experiments generally simplify natural conditions but we used very common species and a soil directly from the region. The main point of our experiment was to show how the same treatment works in different winter climatic settings. We view this aspect as an important improvement to experimental design in climate change studies, as most comparable studies completely depend on single-site conditions. Our approach is still limited to two sites, yet these two sites provide the opportunities to evaluate if treatment effects differ among sites and then would not be of general nature. This can only be done with a completely controlled setup, as otherwise it would be unclear if local climate or rather local species composition or soil conditions would be responsible for observed differences. We agree that such an approach is artificial, yet we believe that such artificial experiments are necessary. We

emphasized the limitations of such an artificial setup in the revised version. We further used the detailed comments provided below to improve the structure of introduction, methods and discussion.

44. A particular concern was the potential that there is no connection between the artificial soil plant environment and the natural systems that the authors are attempting to infer responses to warming for. Further, there was no analysis of the relative weight or interactions among the different parameters for N cycling that were measured.
Response: We have used comparable setups in the past and found that within short time the soil biotic communities reflected the expected differences among plant communities even on the same soil substrate. We added that the sand quarry from which the substrate stems from harbours all studied species in a mosaic of very light pine forests (the dwarf-shrub community species) and meadows (the grassland community species). With regard to the missing analysis of relative weight and interaction please compare to point 31.

45. Is the soil community of this experiment similar to the sites where they are placed?
Response: see above.

46. Do you think that there could be site-specific adaption occurring for the plants and soil microbes found in the colder and warmer areas that are not captured in the mesocosm? For example, is the duration of the experiment (6 weeks warming, planting in May of that year) sufficient to capture plant and/or soil microbial adaptations to different warming/climatic regimes in a way that can add insight to projections of ecosystem responses to warming and change freeze-thaw cycles in these sites? Additionally, it was not clear to me how one would separate the direct effects of warming from the effects of altered freeze-thaw cycles in this study.
Response: Plant communities, soil biotic communities, and soil conditions differ naturally among the two sites. Yet this is not the focus of this study. Here we ask if the same mini-ecosystems (i.e. mesocosms) differ in their response to the warming pulses treatment if they are placed in different climatic conditions. This is one experiment, another future experiment might test the importance of the native communities. Yet we can only do one step at one point and we chose to test the impact of different climates. Due to the setup, it is not possible to disentangle the effects of increased temperature variability and changed FTC. However, we were quite lucky as the FTC were merely changed (±1) and we can therefore link the observed effects to temperature variability (without being able to conclude that FTC would not matter). Anyhow, we feel it quite important to show that temporal variability can also considerably affect N cycling.

47. Are you sure the bait traps are an appropriate technique for this system? How does potential extracellular enzyme activity correlate with or relate to soil biotic activity based on the bait traps? For the PRS Probes, why were they removed more than 2 weeks after the last warming trial? Did you test for saturation of the PRS probes (once saturated, they will no longer accurately measure plant-available N)?
Response: Based on comments by reviewer 4 below, we added a short rational for each response parameter. Bait traps are an easy to use, fast and cheap method accessing decomposition and have been used in these systems (Kratz 1998). Since bait sticks can be eaten by different soil invertebrates and decomposed by soil microbes we would not expect a close correlation with PEEA. The PRS-Probes could not been removed before since the upland site was not reachable due to snow cover. For PRS- Probes we used a test system from western AG innovations Inc. who states in their documentation that at least 12 weeks burial time is okay during winter. The maximum ion capacity for nitrate is given with 2088 µg 10

cm-2. Our values are far lower, so saturation was not a problem here. We included this information in the respective method section now.

48. The discussion has a lot of points that hinge on frost-driven plant damage, but there's a paucity of actually testing to see whether the mesocosm plants were damaged.
Further, there was no effect of the treatment on freeze-thaw cycles. What are the methods by which aboveground biomass was measured (as mentioned in the discussion)?
Response: In hindsight, we would have quantified frost damage! Yet we still feel that all results together only make sense of the plants got damaged somehow. Based on this and other comments above, however, we see the need for an improved flow of argumentation and a clearer separation into facts from the data and interpretation based on reasoning. The biomass data is meanwhile published elsewhere and we refer to this publication: For grassland species complete above-ground harvests were done and for heathland species biometric calculations were done. For further information see Schuerings et al (2014) Winter warming pulses differently affect plant performance in temperate heathland and grassland communities; Ecol Res; DOI 10.1007/s11284-014-1174-x.

Minor comments
49. pg 7799 ln 11-15: It seems like the effectiveness of overwinter plant N uptake remains unclear, rather than now being "fully clear."
Response: In line 11 we exactly wrote that effectiveness of N uptake from plants during winter is not fully clear yet.

50. In the introduction, when ideas about the potential for freeze/thaw cycles and plant species to affect the N cycle were discussed, I felt that the description of what exactly the consequences are for these factors on N biogeochemical dynamics were very vague. For example, in what ways, specifically, does lysing microbial cells affect N cycling? Why does increased overwinter N availability increase the risk of plant frost damage? Why do different plant functional types respond different to winter temperature variability? Although I appreciated the brevity of the introduction, I do feel that these points need to be fleshed out in order to justify the experimental design and really flesh out the hypotheses.
Response: We have improved the description of processes wherever possible, yet we also have to mention that this is not yet possible for all parts because physiological understanding is not yet available for some. Lysing of microbial cells increases the easily available N-pool (we included this information), uptake of N requires physiological activity which reduces frost hardiness because of respiration of carbohydrates which are used for freezing avoidance. The reason for different reactions among plant functional types is unclear.

51. What is the justification/rationale for measuring extracellular enzyme activity? This comes out of left field, especially since no N-targeting extracellular enzymes were assayed. If PEEA was measured, why not look at the enzyme activity changes in ratios?
Regarding the methodology, why run the assays at 20 C?
Response: We use PEEA as an indicator of potential soil biotic activity and state this now (see lines 169-179). Any reduction should indicate some damage to the related organisms.

52. In the discussion, the change in activity as determined by the baits is linked as similar to increases in soil respiration with warming. How do the baits and soil respiration actually compare??
Response: They are not directly comparably since soil respiration accounts for every oxygen using organism whereas the bait test only accounts for decomposing organisms but both can be a proxy for general soil biotic activity.

53. Are you really in a field site at 11 degrees north?
Response: We corrected this mistake.

54. Fig 1: Why is the scale for soil temperature all the way to 20C? It would be easier to read the figure if scaled correctly.
Response: Okay, scaling was improved now.

Anonymous Referee #4 Received and published: 25 June 2014

55. This paper aims to examine the effects of winter climate change and plant species composition on nitrogen cycling in temperate ecosystems. Overall, it is an interesting paper, but there are some major concerns and areas for improvement. While this is a very interesting experiment and the set-up was done well, I do not think the authors measured the appropriate response variables to draw the conclusions that they do.
Response: Thank you for the positive evaluation of focus and setup: In our revised version we used the comments from the referees to improve and clarify our setup, message and interpretations.

56. Throughout the manuscript, I suggest replacing “extreme” with “variable”. It is true that the warming pulses lead to more variation, but I do not consider the resulting soil temperatures to be ‘extreme’.
Response: We don’t use the term “extreme” anymore in the revision. Compare to point 18 and the start of this cover letter.

57. The methods used to examine soil biotic activity are surprising. The authors explain that they use baits, but they do not explain what/who is potentially eating this bait. How widespread is this method? Where has it been used before?
Response: Baits are eaten/decomposed by soil invertebrates and soil microorganisms and are used in many different systems. We included this information in the respective method section now. Bait sticks have been used in grasslands, forests and agroecosystems. For further information see Kratz (1994): The Bait-Lamina Test. General aspects, Applications and Perspectives. ESPR 5 (2), 94-96.

58. It is unclear why if the authors were most interesting in nitrogen cycling, why they didn’t measure mineralization and nitrification directly.
Response: More direct assessments of the N-cycling would have been useful. However, direct measurements of nitrification would have been possible only with high temporal resolution which was not possible during winter. A high temporal replication would also severely disturb the small mesocosms. Net mineralization was planned to be quantified, yet unfortunately the measurement failed due to a technical accident.

59. I do not agree with the authors that they can use atom %15N values as a measure of plant uptake or soil N retention. As the authors explain in the discussion, the plants grew bigger in the warming treatments, which would dilute any 15N taken up by the plants. The authors found lower atom%15N values in the plants of the warmed plots. The plants in both the reference and warmed pots could have taken up the same amount of 15N, but the resulting atom%15N values of the warmed plots would appear lower since they grew more. The results would be much stronger if the authors reported %15N recovery (atom%15N * %N * biomass of plants) in order to get a measure of actual uptake of the 15N tracer.

Response: We are aware of the problem with interpreting the ^{15}N-results. However, we had to make a compromise here. Either we could have harvested the biomass when sampling for ^{15}N at the end of the treatment, or we could studied the longer term response (compare to Schuerings et al. 2014 EcolRes). Yet there is quite some information in the presented data: We find reduced atom%^{15}N and stable or reduced biomass production (again see Schuerings et al. 2014 EcoRes) in the warming pulses treatment. The proposed dilution effect can therefore be rejected. While we cannot quantify the reduced N-uptake absolutely, our data still shows that it was reduced. We clarified this point in the discussion of the revised version (ll. 339-341).

60. The authors use the term "N cycling" throughout the manuscript, but they are actually using proxies for N cycling (e.g., atom%15N in plants, baits eaten, etc.). It's unclear why the authors did not measure soil N cycling directly (i.e. mineralization and nitrification) or plant N uptake (as %15N recovery). It is true that they measure soil N pools, but these are pools and not fluxes.
Response: See above, we have now formulated more carefully.

61. I suggest that the authors restructure the Introduction to more clearly set up their work. They created 6 warming pulses in winter and then measure 15N in plants within winter. Are they expecting N uptake by plants in winter in these ecosystems? Does most precipitation in winter come as snow or rain in this system?
Response: N uptake by plants during winter has commonly been reported (see introduction), also for our species. So yes, we expected winter N uptake. We clarified this aspect. At the lowland site most precipitation comes as rain but at the upland site it is snow, this is one of the reasons for testing the response at these two sites.

62. Page 7798, Line 21: Change "plant performance" to "plant nitrogen uptake" to be more specific.
Response: Changed as suggested.

63. Page 7799, Line 17: Katherine Hayhoe's downscaled model projections for snow cover should be cited here.
Response: We considered this reference but found only model projections for American regions. So we left this out.

64. Line 23: Insert ": : :"can physically damage plant roots (Tierney et al. 2001) AND REDUCE THE ABILITY FOR PLANTS TO TAKE UP NITROGEN (CAMPBELL ET AL., IN PRESS GLOBAL CHANGE BIOLOGY). Detailed reference: Campbell et al. in press, Global Change Biology. Increased nitrogen leaching following soil freezing is due to decreased root uptake in a northern hardwood forest.
Response: We included the reference and changed the text as suggested.

65. Page 7800, Line 13: "we tested the effects OF more extreme: : :"
Response: We corrected this mistake.

66. Page 7802, Line 12: Why was 1900W chosen for the treatment? Does this match with model projections for this area?
Response: This warming treatment was administered to increase variability of soil and air temperatures and therefore decrease snow cover. Since only warming lamps weren't always enough to melt the snow cover depending on weather conditions (wind, fog, etc.). The resulting mean warming is in the projections for this region. There is no trustworthy scenario

on temperature variability or temperature extremes (Compare to Kreyling, Jentsch, Beier 2013 Ecology Letters).

67. Page 7802, Line 15: Does this refer to soil or air temperature?
Response: We clarified this.

68. Page 7802, Line 17: This text implies that soil temperature was only measured in the treatment plots, but the figure suggests soil temperature was also measured in the reference plots. I suggest making this clearer.
Response: There a mistake has happened. Soil temperature was measured once in every block (5 treatment blocks and 5 reference blocks per site). We clarified this.

69. Page 7803, Lines 10-17: I am not familiar with this technique and I imagine that many readers will not be. I suggest that the authors explain who could be eating these baits and whether they are microbes, micro or macro arthropods, etc. Also, what is the threshold amount of light absorbance needed to indicate that the bait has been 'eaten'? How biased is this method?
Response: Baits are eaten by soil invertebrates (macro- to micro arthropods, earthworms, etc.) and soil microorganisms. Analysis was done by one person who was blind to the factors. We included this information in the text.

70. Page 7806, Line 20: remove the word "extreme"
Response: Changed to "increased soil temperature variability".

71. Page 7807, lines 4-7: When the authors compare the two sites and compare species, it's unclear if they pooled across all of the plots or the reference ones only. I suggest pooling across the reference only to permit comparisons of site or species only without the interacting effects with warming.
Response: We pooled across all plots here. Statistically, this is correct as we tested for potential interaction effects. We included the term "overall" here.

72. Page 7808, Line 20: The authors did not measure N cycling directly and therefore this term should be removed.
Response: Changed the sentence to: "Recurrent winter warming pulses led to increased soil temperature variability and influenced N cycling in our experiment."

73. Page 7808, Line 19: remove "extreme" from the text.
Response: Done.

74. Page 7810, Line 15: Again, one cannot use atom%15N values as a proxy for "15N incorporation", especially when growth and biomass of the plants under different treatments vary.
Response: Compare to point 59.

75. Page 7811, Line 9: I suggest also citing Campbell et al. Global Change Biology in press here (complete reference information above).
Response: We included the reference there.

76. Table 1: The years are unclear in this table. The legend says "until 2008". What do the years for each row refer to? Why do some go from earlier to later (e.g., 1998/1994) and some go from later to earlier (e.g., 2003/1999)?

Response: Years in brackets give starting year of measurements on site. Measurements started in different years.

77. Figure 1: I suggest making the two lines more distinct and including error bars since there are 5 replicates for each treatment.
Response: Enhancing the contrast between lines is done but we already tried to incorporate error bars but the figure becomes nearly unreadable then.

78. Figure 2: I suggest making the units on the y-axis to be ug N per unit area per day. Are the cores really 10cm2? That seems quite small. Are the values for W vs. C the reference plots only? Same question for the community figures. Figures 3-4: Same comments as above.
Response: In all graphs we pooled across all plots (except Fig. 4 d, stated in the text). Since ion uptake by resin sticks is not a linear process, it is not possible to calculate values for smaller time units.

Anonymous Referee #5 Received and published: 8 July 2014

General comments:
79. The authors have presented results from an interesting experiment investigating the effects of altered winter soil temperature regime on several soil processes, at two different sites and under varying plant species. The subject is important: I personally think too much of our understanding of soil nutrient cycling is based on observations during the warmer months. In this sense the authors are contributing useful information for those of us trying to understand the controls of soil biogeochemical cycling, particularly in the context of climate variability. The paper is concisely and clearly written, and the data and analyses appear mostly sound. However I think the presentation and some aspects of the interpretation of the results could be improved in several areas.
Response: Thank you for this positive evaluation of focus and study design and for the very constructive and detailed critique which helped us to improve the presentation of the manuscript!

Specific comments:
80. My main criticism of the manuscript is that the authors too often blur the distinction between the parameters they are actually manipulating and measuring, and the conceptual parameters that their study is trying to address. For example: the number of eaten baits is referred to as "biotic activity", the relative enrichment of 15N in plant tissues is referred to as "plant N uptake", the accumulation of inorganic N by resin sticks is referred to as "N availability". This kind of shorthand is useful, and we all use it, but I had the sense reading the manuscript that the authors uncritically accepted that their chosen response parameters were unbiased and realistic indices for the processes and states they are interested in. I think the manuscript could be improved in this area by adding two types of additional content: 1) in the Methods section, a bit more time should be spent explaining why each of the chosen response parameters is appropriate for measuring the processes of interest for the hypothesis, 2) and again in the Discussion section, include a discussion of the extent to which biases and limitations inherent in the methods may have influenced the results and subsequent interpretation. Studying soil processes is tricky, and we always use indirect methods such as those in the present study - but I personally think we should be transparent in our discussion of both the rationale for choosing a given indirect method, and the limitations our chosen method place on interpretation.

Response: We adapted our presentation to common standards in comparable papers but we completely agree. We added some rationale and explanation of limitations for each response parameter in the methods and point to this again in the discussion.

81. Manipulating temperature variability while maintaining realistic field conditions is devilishly hard, and the authors are to be commended for their efforts! However, the chosen method (applying heat pulses) means that their chosen variable (temperature variability) is necessarily confounded with another (absolute temperature). Given that many of their chosen response variables can be expected to respond to higher temperatures alone, I would have appreciated a more thorough discussion of how, and to what extent, we can separate variability effects from pure warming effects in the current study.
Response: It is definitely problematic to disentangle the effects of minimum and maximum temperature (i.e. temperature variability). However, the reduced N-uptake by the plants can hardly be explained by increased winter temperature alone (see point 1). Furthermore, for the upland site, mean temperature was only increased by 0.2 C and FTC only changed by 1, therefore we can expect some of the effects happening due to increased temperature variability. We emphasized this aspect in the revision of the discussion and clearly mentioned which of the observed effects could be explained by warming alone and which effects can only be explained by acknowledging temperature variability.

Some more minor comments:
82. page 7798, lines 1-6. There is no real question in this abstract, I suggest the authors fill in the rather large logical leap between the first and second sentences here at the beginning.
Response: We agree and replaced the second sentence by “Here, we experimentally evaluated the effects of increased winter soil temperature variability on selected aspects of the N-cycle in mesocosms containing different plant community compositions.”

83. page 7798, line 11 and elsewhere throughout suggest "potential activity"
Response: We agree that using potential activity throughout the manuscript is more precise and changed this accordingly.

84. page 7798, line 17, as mentioned at point 2, I would need to be more convinced that the observed effects weren’t just due to warming alone before I accept this conclusion.
Response: See point 81.

85. Introduction, 1st paragraph. This entire introductory section applies mainly to temperate zones and colder. I.e. places where there is winter snow cover, and minimal winter plant and soil activity. There are other types of ecosystems (e.g. mediterranean-type, arid lands) where winter is the most active season! Perhaps somewhere near the beginning the authors should specify they are concerned with temperate/boreal/...? systems only.
Response: True, we implicitly talk about temperate and colder systems yet we explicitly mention this now.

86. p. 7800 lines 8-10 – for who or what is this “increased responsiveness” “beneficial” or “detrimental”?
Response: This sentence is about productivity of the plants and this is now included.

87. p. 7800 lines 13 – “extreme” is often a problematic term. Extreme relative to what? and for whom? You could say that winter warming leads to less extreme temperature regimes, as it brings temperatures closer to the year-round average. Perhaps “more variable” is a more accurate term?

Response: We can agree on this, compare to point 18. We don't use the term "extreme" in the revised version.

88. p 7801 line 8 – E and N seem to be swapped here.
Response: Yes, a typesetting mistake happened here which is now corrected.

89. p. 7802 The authors should supply more detail about the warming treatment. How long were the warming pulses? How was "frost" defined (air temp? soil temp? weather station? where?). What was the rationale for choosing to warm only on days when frost was forecasted (by whom?) to continue for at least 48h?
Response: Frost was defined as soil frost measured directly on site. When local weather forecast predicted further air frost for more than 48h (planned length of a warming pulse, realistic scenario for the region), warming was applied. We included further information on this. Limiting the warming pulses to times with natural frost was thought to insure that we actually altered temperature variability with as few effects on maximum temperature as possible.

90. p 7803-7805 – as noted above (point 1) I think it would be helpful to give some justification for the choice of response parameters. Some previous work or rationalization for why the things measured are good indices for the things the authors are interested in. Only a sentence or two in each case, but then the reader has a better idea of the design decisions underlying the study and its interpretation.
Response: We incorporated some rationale for each response parameter in the resubmission.

91. p. 7804 – there is a little too much technical detail in the description of the soil enzyme assay protocols. I suggest referring to the standard protocol publication, and only note the reaction conditions specific to this study (e.g. enzymes assayed, buffer conditions, choice of concentrations and incubation temperature and times, range of calibration curve etc).
Response: We shortened the paragraph. It now says: "For the potential extracellular enzymatic activity (PEEA), which we used as another proxy for soil biotic activity and decomposition, three soil samples (2 cm diameter, 10 cm depth) per mesocosm were collected and mixed for assays of potential extracellular enzyme activity in soil on 21 February 2011. Soil samples were stored in airtight plastic zip-bags at 4°C and were analysed within 3 days. PEEA assays were carried out with Methylumbelliferone substrates (MUF) (Pritsch et al., 2004; Pritsch et al., 2005). The following PEEAs were measured: MU-β-D-glucopyranoside (MU-G), for β-glucosidase, MU-β-cellobioside (MU-C) for cellobiohydrolase, MU-β-D-xylopyranoside (MU-X) for xylosidase, MU-phosphate (MU-P) for acid phosphatase. Substrates and calibration saturation and incubation times were determined in pre-experiments (data not shown) as follows: MU-G and MU-X each 500 µM incubating for 60 min, MU-C 500 µM incubating for 120 min, MU-P 800 µM incubating for 40 min. Fluorescence was detected at an excitation wavelength of 360 nm and an emission wavelength of 450 nm with a Gemini EM Fluorescence Microplate Reader from Molecular Device, California."

92. p. 7806 line 20. suggest "larger variability in soil temperature"
Response: see point 87

93. p. 7808, line 3. How was this calculated? If this is based on different concentrations of label in plant tissues then the calculations should also include any differences in plant biomass due to treatments to truly reflect "uptake and incorporation".
Response: compare to point 59.

94. p. 7808, lines 22-25 and p. 7809 line 6. Increased FTC frequency doesn't seem a very likely explanation for the patterns seen here, as apparently FTC was not affected by the treatments? p. 7807 lines 2-3.
Response: Yes, that is our message: despite only small changes in FTC we found significant effects. In the cited publications FTC are used as an explanation. We show that besides changed FTC regimes, temperature variability is also an important aspect to be considered.

95. p. 7809 line 11. "Freezing intensity is ... therefore another important determinant of N cycling response" – I'm not sure the data support such a strong assertion! How can we be sure that it is the lower minimum in warming-pulse-treated plots that is responsible for the N effects? Higher overall temperature seems a more parsimonious explanation (see point 2 above).
Response: "Freezing intensity" does indeed not describe our main message well. We now added a half-sentence that freezing intensity is important for plant frost damage which indirectly can influence N cycling. Compare to point 81 for more detail.

96. p. 7810 lines 4-19. Given that warming led to higher "availability" but less "uptake" of N, what can be said about the relative importance of supply (from micro-organisms) vs demand (by plants) in controlling the response of N dynamics to the applied treatments? In other words, are the observed changes in N availability driven by increased mineralization by microbes? or decreased uptake by plants? or both?
Response: A very good question. Our data implies both, reduced plant N uptake and increased N mobilization. Unfortunately, we cannot answer which of the two processes are more important, yet we can show that both processes are acting.

97. p 7810, line 25. I may not have followed the logic properly, but it seems like the treatment led to less N "uptake" without a change in the biomass of the plants. For this to happen either the plants have a drastically different C:N ratio, or they are accessing alternative sources of soil N. I think this complicates the authors' interpretation that differences in atom % 15N can be straightforwardly interpreted as difference in total N uptake.
Response: See point 59 above.