We thank the reviewer for their comments, and helpful ideas for improving the manuscript. Please find below our point-by-point response, and below this the proposed revised manuscript and additional supplementary material follows (the latter is titled 'Sensitivity to snow' and will be included with the supplementary material from the original submission).

Model description
How about a Table summing up the 3 models main features? This would allow significant text shortening. Also please make sure equivalent information is given for all 3 models. For example, vegetation details are lacking for JSBACH.
The model description section is now amalgamated into a single section, so that anything that is common to all 3 models is only mentioned once, and differences between models are explicitly pointed out. This way equivalent information is now given for each one. This has shortened the text. we have also added a table summarising the main model features, for clarity – see Table 1 in the marked up manuscript, following these comments.

By the way, PFT is defined nowhere and some institute abbreviations are not explained (IPSL, NCSDC). I let the editor decide whether that is necessary.
We have defined PFT and NCSCD. We have left the abbreviations for the Earth System Models as their full names are not relevant, but the editors can request that to be changed if necessary.

Lichens are not mentioned in any model description, from which I assume that they are not considered. Yet, they can be very abundant at some Arctic sites, sometimes covering most of the ground. They have physical and biological properties very different from mosses, for example a much lower thermal conductivity and different hydrological properties which strongly impacts the ground thermal and hydrological regimes. Please consider specifically mentioning this omission.
The moss model in JSBACH is actually designed to represent 'average' properties of bryophytes and lichens, rather than considering them separately. One reason for this is that the properties can vary more between species than they do between the phyla as a whole. In the model description we have added the following: "This model represents both mosses and lichens by one plant functional type with 'average' physiological properties." However as you suggest, it can be important to consider them separately and we have added a comment on this in the discussion (see below).

A couple of sentences or a line in the future model Table to describe the snow scheme would be nice (single layer, multilayer...). In fact all the models have a multilayer snow scheme. We included this in the re-written model description section and in the new table (see above).
Please also specify here that nutrient aspects are not treated in any of the 3 models. Done

Site description
The description of all sites should really be homogenized and considerably shortened, by at least 3 pages. What is in Table 1 need not be repeated in the text. Incomparable data are often given in the text. For example, some sites have mean annual temperature, others January and July, please be consistent.
We have shortened these to less than half a page for each site, and removed everything from the text that can also be found in the table. See marked-up manuscript following these comments.
Also detail the snow fraction of precipitation in all cases: "most" is vague and not very useful.
Snow fraction of precipitation can vary a lot from year to year at some of these sites, but nonetheless we have found some indicative values in the literature and added these to the table of climatic and permafrost variables.

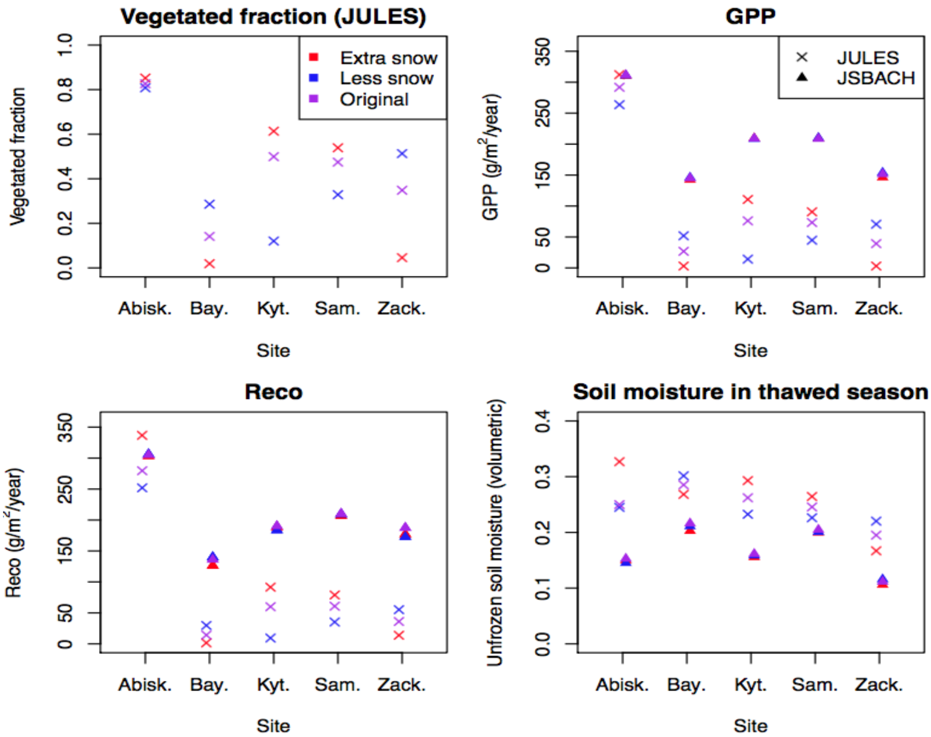
All plant Latin names must be in italics. By the way, line 219, what are the Salix? Richardsonii, arctica, other? The Salix are usually Salix pulchra, which is a small (up to 50 cm high) willow

shrub. We have checked all of the latin names in the revised version.
Field data
Measuring snow precipitation and snow depth in a reliable and representative manner is always a problem in the Arctic and the text does not convince me that this aspect was treated properly. Moreover, its impact may be understressed here since it conditions the permafrost thermal regime and therefore all carbon processes. How about details of the precipitation measurement, such as the presence of a wind shield around the gauge? I understand that precipitation measurements were not used, but since snow depth measurements are not convincing, as detailed below, perhaps analyzing precipitation data in more detail would be useful.
First of all, while precipitation measurements were not used for snowfall, they were used for rainfall, for which they are more reliable. This was not made clear in the manuscript so we have added a note: "However, the local precipitation measurements were still used for rainfall, as this is much more reliable, with an average undercatch of around 10% (Yang et al., 2005)."
At some sites there is no wind shield (e.g. Samoylov and Bayelva), and at others there is a wind shield (e.g. Abisko). We have added after the above line "(depending on the set-up of the precipitation gauge, which differs between sites)"
For snowfall, as discussed, the direct precipitation measurements are not reliable. However, to address the question about the impacts of the uncertainty in snowfall forcing, we have performed a sensitivity study and assessed its impact on the carbon-cycle processes – see below.

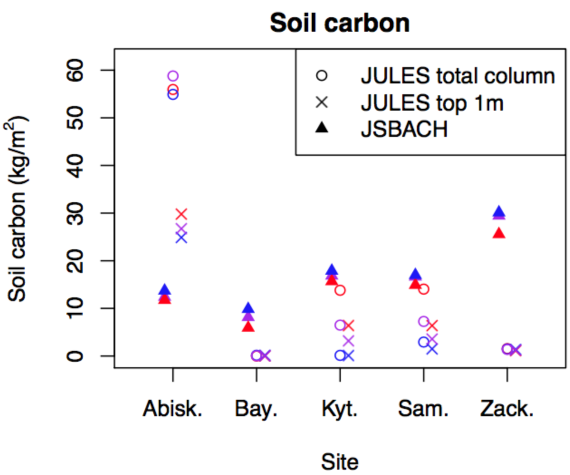
Was there any attempt to correct measured snow precipitation as described in (Forland et al., 1996)? This can double estimates of precipitation amounts and considerably improve agreement with snow accumulation.
Thanks for this suggestion. At several of these sites, the precipitation gauges installed do not actually detect snowfall, reducing the possibility to apply this in this study – along with the issue of wind-redistribution that reduces the correlation between precipitation and snow depth on the ground (for example, we found no correlation between snow depths at the Abisko mire and the nearby research station). However it is certainly a good idea to consider this approach for future studies. In general, for this study, we considered using the observed snow depths to be the best way of constraining snow precipitation for these sites, but additionally we have now performed a sensitivity study (see below) to show the impact of the potentially large uncertainties in snowfall forcing.

Measuring snow depth in a representative manner is difficult. Certainly using one point measurement is inadequate. In particular, in low-centered polygons, variations are huge and at least 100 measurements are required for a representative value. Please detail the representativity of your snow depth measurements. In case the data are found to have limited representativity, this should be clearly stated and perhaps a sensitivity study would be useful (if it is still possible to perform it): what is the impact of snow amount on permafrost temperature and carbon cycling?
The snow depth measurements are point measurements for most sites, except for Abisko, where measurements are averaged from several locations on the mire. Considering the representativity for each site: The Abisko measurements are deliberately taken to give a representative sample. At Zackenberg, there is a CALM grid at the site where snow depths are measured periodically. Snow depths are relatively homogeneous here and the point observation appears to be representative. At Samoylov and Kytalyk there will be variability due to polygon structures and wind distribution as you suggested, so in general point observations are not representative (see also the comparison with GlobSnow snow water equivalent, below – here it appears that the Samoylov simulation matches better with the GlobSnow product than Kytalyk.) Finally at Bayelva, it is hilly and there can be some variation in the flux tower footprint, so the point observation may also not be representative for this site, in fact it seems to be a little higher than the 'typical' values. We have added a full discussion of the representativity of snow depths in the supplementary material. We have then performed a sensitivity study with two of the models for all of the sites, which aims to cover all uncertainties including where the single snow depth measurement was not deemed representative of

the flux tower area. The runs were repeated twice with snowfall increased by 50% and reduced by 50%, respectively. We added the details of this in the supplement and some discussion in the text. In general there can be significant differences in the carbon cycling, in particular for JULES – this is because the snow impacts the soil moisture availability. For two of the sites (Kytalyk and Samoylov) this resulted in very different vegetation fractions during spinup and therefore a big difference in soil carbon stores. For JSBACH, however, the differences are fairly minimal. It is clear that the differences in GPP and Reco are due to soil moisture in JULES as the vegetation only responds to soil moisture and climate forcing in the model, and we see clearly the same patterns in all these variables:



As expected, all sites show an overall warming of the soil due to increase in snow depth, with the majority of the warming in winter. In JSBACH this can also be seen to impact the soil carbon stocks (following figure). In JULES, the impact of vegetation differences on soil carbon is larger than the impact of warming and dominates the changes.



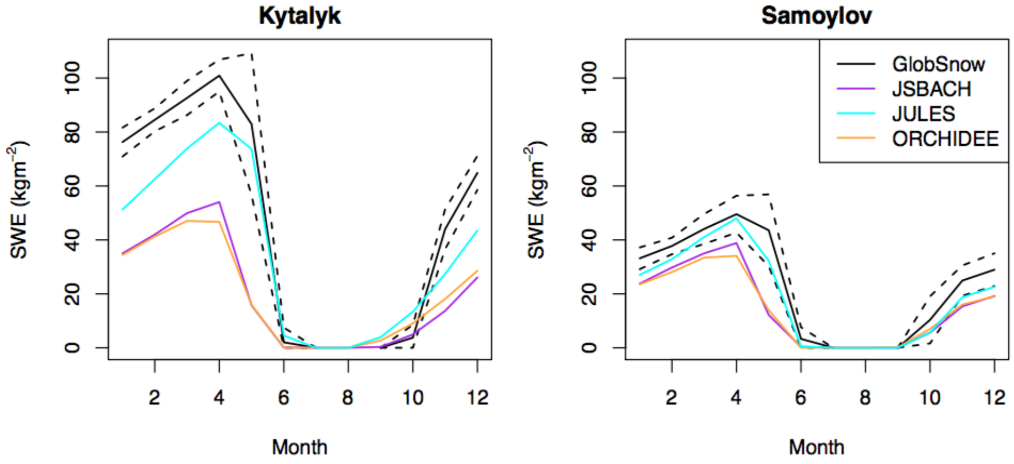
A full discussion of the sensitivity study, with plots, is added in the supplementary material (also included following these comments). In the main text we have added some discussion in the section on snow, regarding the poor simulation for Abisko (see below), and the following: “It is important to be careful when modelling snow depth based on single point observations, as they may not be representative of the area as a whole. Further details on the representativity of snow depths are given in supplementary information. The sensitivity of carbon cycle processes to increased/reduced snowfall is discussed in Sections 3.5 and 3.6.1."

We then include discussion in Sections 3.6.1 (see below) and 3.5: "The soil carbon stocks are sensitive to changes in snow depth in these models (see supplementary Figure S8), through changes in soil temperature (JSBACH) and changes in vegetation growth (JULES). In JULES, both vegetation and soil temperature changes affect the soil carbon, but the vegetation effect dominates. In fact, for two of the sites (Kytalyk and Samoylov), the vegetation coverage is so different during spinup that the simulation with increased snowfall accumulates twice as much soil carbon as the default case (although the stocks are still much too small and the absolute difference is less than 10 kgm^{-2} in the whole soil column)."

A note about the sensitivity study is also added in the methods: "Even with these corrections, there is still considerable uncertainty in precipitation forcing, particularly the snowfall, so in order to test the impact of this, two of the models (JULES and JSBACH) performed two additional sets of simulations, with snowfall increased and reduced by 50%."

Perhaps looking at data from reanalyses would also be helpful for an extra evaluation of precipitation and snow depth data.

Since the snow depth is not controlled by precipitation at many of these sites, but much more by the wind, we decided not to look at any more precipitation data. However, the Globsnow reanalysis data could be useful to compare against SWE in the models. Unfortunately, for Zackenberg (Greenland) and Bayelva (Svalbard), there are no values in the dataset as these are very small pieces of land between glaciers and ocean. There is also no value given for the closest pixel to Abisko (which may be because the site is next to a lake?). Taking the next pixel along gives an SWE that is much higher than could be expected for the mire site, which is not surprising given the landscape is mountainous and snow depth will be very variable around this region – however, this precludes using Globsnow data for Abisko. This leaves two sites: Kytalyk and Samoylov, which are flatter and more homogeneous landscapes where the product should be more representative: Langer et al. (2013) showed that the globsnow SWE data matched well with the Samoylov island data assuming constant snow density of 250 kg/m^3. We can compare directly with modelled SWE. Since we can only do this for 2 out of 5 sites this does not merit an extra figure in the text but we include a plot here (showing average of 2005-2013):



This shows a reasonable simulation of SWE for Samoylov but too little for Kytalyk (despite the models matching snow depth quite closely). This may be because there is a larger uncertainty in Kytalyk snow depth, due to having a limited number of years of in situ measurements, or because the snow is more compacted in reality than the models, or alternatively because the point measurement is not representative of the larger area. The GlobSnow product also varies in accuracy depending on proximity to ground stations. Further investigation would be required to confirm the reason for the discrepancy.

Are there any field measurements of snow density to validate model assumptions of this variable?
There are field measurements or literature values available from some of the sites. We can also output this from the models. In JSBACH the snow density does not vary much between sites, whereas for JULES and ORCHIDEE, density varies more between sites, but is quite consistent between these two models, suggesting that they are constructed similarly. At Samoylov, the estimated density is between 200 and 400 kgm^{-3} (20th April) depending on the type of snow (Gouttevin et al., 2017), whereas JULES and ORCHIDEE simulate a lower density (around 180 kgm^{-3}). The work of Gouttevin et al. (2017) suggests that the reason for this difference is likely because the models do not simulate wind compaction. Similarly at Zackenberg, the average density for April-May is high (around 375 kgm^{-3}, https://data.g-e-m.dk), and the models simulate a lower density (270-350 kgm^{-3}). On the other hand, at Bayelva the mid-season snow density is 305 kgm^{-3} (Gisnås et al., 2014), which is very close to the values simulated in JULES and ORCHIDEE. In the Supplementary material we have added a comment on this: "It is also useful to compare snow density in models and observations. For example, recent work shows that including wind compaction is essential to capture high snow density at Samoylov (Gouttevin et al., 2017), and indeed our models show a snow density closer to the 'default' models in Gouttevin et al. (2017), which is too low due to the omission of wind compaction processes."
It would make an interesting study focussing on the snow dynamics in these models at these sites – we hope that by collating all of the data for this study we have opened up opportunities for further detailed studies. We have added a comment on this at the end of the conclusion: "This work also opens up opportunities for further process studies in future."

By the way, snow temperature measurements at several heights can be very useful to evaluate the validity of snow schemes, and implementing those at the sites described here may be valuable for future work (Barrere et al., 2017).
We have added a comment on this in the supplementary material: "It is also important for the models to better represent the profile of snow thermal conductivity: for example the models do not simulate the low-conductivity 'depth-hoar' layer that can form at the base of the snowpack (Domine, et al. 2016). For this, monitoring of snow temperature at different heights can be valuable to improve the models (Barrere et al., 2017)."
Along with further comments on the need for better representation of snow in these models.

Lines 327 and 329: please use "snow depth" throughout.
Done

Results and discussion
Lines 431-432. The snow depth model output at Abisko is not "reasonable". It just does not seem to work there. Please consider representativity of field data and modify discussion.
The snow depth measurements for Abisko are actually an average taken from several locations on the mire and are representative for snow depth on the mire. The representativity of the snow data is now discussed in detail in the supplementary material. A more realistic simulation for Abisko has now been made in the -50% snow case, in our snow sensitivity study. We showed that, in general, the models are sensitive to changes in snow. However at Abisko the carbon stocks/fluxes do not show major changes. We have added a comment in the snow results/discussion section: "...for the most part the models make a reasonable simulation of the snowpack accumulation and compaction, with the exception of Abisko where the models are all biased high. Here, snow inputs are particularly uncertain as no high-resolution timeseries of snow depth are available (unlike the other sites). We performed a sensitivity study to test the impact of uncertainties or variability in snow depth on the simulated carbon-cycle processes. In this study, a reduction of 50% in snowfall allows the models to simulate a realistic snow depth at Abisko – see supplementary material. The impacts on soil carbon stocks and fluxes are fairly small, however (between 0.2% and 10%, supplementary Figures S7 and S8)."

Line 433. “snow often melting a little too early” in simulations. Ambiguous as written.
This sentence is re-written as: “During the melting season the models are less accurate than during accumulation, with the snow often melting too early - by up to 25 days in the most extreme case.”

Lines 436-437. How do models account for vegetation effects on snow albedo?
Snow albedo is reduced by the presence of vegetation, more so when the snow is shallower or the vegetation is taller. (For example, in JULES, the albedo is interpolated between the snow albedo and the snow-free albedo according to snow depth, d, and the vegetation roughness length, z_o (Essery et al 2003): snow-covered fraction = $d / (d + 10z_o)$)
In the text we have added: “(this is modelled by interpolating between snow-covered and snow-free albedo depending on snow depth and vegetation height)”

Lines 457-458. How about thermal conductivity values obtained by JULES, and how do they compare with other models? Perhaps also compare with values obtained at a comparable high Arctic sites in low-centered polygons (Domine et al., 2016) if you think this supports your case. Note by the way that stratification of thermal conductivity can have an important effect, as suggested by (Barrere et al., 2017), so that one-layer snow models can give the correct mean thermal conductivity value while making a large error on atmosphere-ground heat fluxes. Incorrect snow thermal conductivity stratification can also lead to incorrect timing of ground freezing and thawing. Arctic snow often has a very low thermal conductivity layer at the base, which delays freezing and thawing. This process is missed if the snow scheme gives a high thermal conductivity to the basal snow layer.
Thanks for this, it was helpful to compare the values in JULES with these observations. We have added in the main text: “Indeed, the conductivity of snow in the JULES simulations is between 0.03-0.1 $Wm^{-1}K^{-1}$ at the sites with shallow snow (and in the upper layers of the snowpack at sites with deeper snow), which is considerably lower than typical values for similar tundra sites, which suggest a realistic conductivity would be around 0.2-0.3 $Wm^{-1}K^{-1}$, at least for the upper part of the snowpack (Gouttevin et al., 2012b; Domine et al., 2016).”
We have also added a comment about the need to represent a low conductivity layer at the base of the snow in the supplementary discussion (see above).

Line 559. A word on nutrients here?
We have added this, so this part now reads:
“Of these, climate is the main driver of vegetation growth in these models (since nutrient limitation is not included, the soil only impacts the vegetation through moisture stress...)”

Line 574-575. “GPP depends mostly [...]on shortwave radiation in the second half of the season”. How about moisture? For example, (Frost and Epstein, 2014) stated that “rates of shrub [...] expansion were not strongly correlated with temperature trends and were better correlated with mean annual precipitation”.
This is a good point, vegetation growth is correlated with soil moisture, and our sensitivity study with increased/decreased snow depth has confirmed this in the JULES model. However (as your quote implies), the moisture effect occurs over longer timescales than daily and hourly variability, which is driven more by shortwave radiation. The sentence in the manuscript will be better phrased as “In particular, the increase in GPP in the first half of the season is driven by increasing LAI, and the downward trend of GPP in the second half of the season is driven by shortwave radiation”.
We have then added a comment about the impacts of soil moisture (end of Section 3.6.1):
“Carbon fluxes are also sensitive to soil moisture, as seen in simulations with increased/decreased snowfall, where differences in soil moisture availability in summer are reflected by changes in annual mean GPP, ecosystem respiration and vegetation fraction in JULES (Supplementary Figure S7), in line with Frost and Epstein (2014). Therefore, realistic simulation of precipitation and soil

moisture is a pre-requisite for improved LAI and vegetation dynamics."
We have also added a sentence in the conclusion: "There is also a need to address remaining issues in the model physics, particularly for soil moisture and snow."

Conclusion
The impact of mosses is stressed, but as mentioned above, I really think that lichens can have a huge impact. I gather that they are not very important at the sites studied here, but on a pan-Arctic scale, this is probably different.
We agree, and we have added in the discussion: "It could also be important to consider lichens separately from mosses, as their physical and biological properties can be very different. For example, the high albedo of lichens can impact the Earth's radiation budget (Bernier et al., 2011)."

Since you are talking about landscape dynamics, you may talk about the impact of lakes and ponds caused by landscape dynamics such as thermokarst lakes formation. These lakes are often hotspots of GHG emissions. See e.g. (Bouchard et al., 2015) and references therein.
We agree this is a significant omission from our models on a large scale. We have added a note on this in the conclusion: "Lakes and ponds also play a major role in methane and carbon dioxide exchange with the atmosphere (Bouchard et al., 2015; Langer et al., 2015) and should also be considered in future land surface models."

Figure1. What is the meaning of mean snow depth? Spatial mean? Temporal mean? Over what period? The Abisko graph does not seem to match the mean value.
The plots show 'mean annual cycle', and this means we take the average for a given time of year across a number of years observations – but generally for a single site. The years used for every site and for every variable (including snow) are given in the supplementary information and this is referenced in the captions for all the figures. In fact, Abisko is slightly different from the other sites in that the observations have very low temporal resolution, but they are in fact averaged across different locations on the mire and therefore are representative for the area as a whole. On the figure caption, we have clarified: "Mean annual cycle is calculated from a single site over a number of years, except for Abisko where measurements were taken in several different locations on the mire."
Regarding the mis-match between the models and observations at Abisko, we have added some discussion on this – see above.

Table 1: What is summer? What is winter? Permafrost T, at what depth?
Permafrost temperature is not measured at a consistent depth at the sites, these are approximate values anywhere below the active layer but they give an indication of the differences in permafrost conditions between sites. Summer and winter generally refer to the maximum and minimum average monthly temperatures. We have clarified this in the table by changing "summer" to "max. monthly" and "winter" to "min monthly".

References

Essery, R.L., et al.: Explicit Representation of Subgrid Heterogeneity in a GCM Land Surface Scheme. *J. Hydrometeor.*, **4**, 530 543 (2003)
Gisnås, K., et al.: A statistical approach to represent small-scale variability of permafrost temperatures due to snow cover, *The Cryosphere*, **8**, 2063-2074, https://doi.org/10.5194/tc-8-2063-2014 (2014)

Gouttevin, et al.: Observation and modeling of snow at a polygonal tundra permafrost site: thermal implications and spatial variability. Submitted to *The Cryosphere* (2017).

Langer, M., et al.: Satellite-based modeling of permafrost temperatures in a tundra lowland landscape, *Remote Sensing of Environment*, **135**, 12-24 (2013)

Manuscript prepared for Biogeosciences
with version 2014/09/16 7.15 Copernicus papers of the LaTeX class copernicus.cls.
Date: 9 August 2017

Carbon stocks and fluxes in the high latitudes: Using site-level data to evaluate Earth system models

Sarah Chadburn[1,2], Gerhard Krinner[3], Philipp Porada[4,5], Annett Bartsch[6,7], Christian Beer[4,5], Luca Belelli Marchesini[8,9], Julia Boike[10], Bo Elberling[11], Thomas Friborg[11], Gustaf Hugelius[12], Margareta Johansson[13], Peter Kuhry[12], Lars Kutzbach[14], Moritz Langer[10], Magnus Lund[15], Frans-Jan Parmentier[16], Shushi Peng[3,17], Ko Van Huissteden[9], Tao Wang[18], Sebastian Westermann[19], Dan Zhu[20], and Eleanor Burke[21]

[1]University of Leeds, School of Earth and Environment, Leeds LS2 9JT, U.K.
[2]University of Exeter, College of Engineering, Mathematics and Physical sciences, Exeter EX4 4QF, U.K.
[3]Laboratoire de Glaciologie et Géophysique de l'Environnement (LGGE), 38041 Grenoble, France
[4]Department of Environmental Science and Analytical Chemistry, Stockholm University, 10691 Stockholm, Sweden
[5]Bolin Centre for Climate Research, Stockholm University
[6]Vienna University of Technology, Vienna, Austria
[7]Austrian Polar Research Institute, Vienna, Austria
[8]School of Natural Sciences, Far Eastern Federal University, Vladivostok (Russia)
[9]Department of Earth Sciences, Vrije Universiteit (VU) Amsterdam, The Netherlands
[10]Alfred Wegener Institute, Helmholtz Center for Polar and Marine Research (AWI) 14473 Potsdam, Germany
[11]Center for Permafrost (CENPERM), Department of Geosciences and Natural Resource Management, University of Copenhagen, Denmark
[12]Department of Physical Geography, Stockholm University, 10691 Stockholm, Sweden
[13]Dept. of Physical Geography and Ecosystem, Lund University, Sölvegatan 12, 223 62 Lund, Sweden
[14]Institute of Soil Science, Center for Earth System Research and Sustainability, Universität Hamburg, Hamburg, Germany
[15]Department of Bioscience, Arctic Research Center, Aarhus University, Frederiksborgvej 399, DK-4000 Roskilde, Denmark
[16]Department of Arctic and Marine Biology, UiT - The Arctic University of Norway, Tromsø, Norway
[17]Sino-French Institute for Earth System Science, College of Urban and Environmental Sciences, Peking University, Beijing 100871, China
[18]Key Laboratory of Alpine Ecology and Biodiversity, Institute of Tibetan Plateau Research and Center for Excellence in Tibetan Plateau Earth Sciences, Chinese Academy of Sciences, Beijing 100085, China
[19]Univesity of Oslo, Department of Geosciences, P.O. Box 1047 Blindern, NO-0316 Oslo, Norway
[20]Laboratoire des Sciences du Climat et de l'Environnement, LSCE CEA CNRS UVSQ, Gif Sur Yvette, France
[21]Met Office Hadley Centre, Fitzroy Road, Exeter EX1 3PB, U.K.

Correspondence to: Sarah Chadburn (s.e.chadburn@exeter.ac.uk)

Abstract.

It is important that climate models can accurately simulate the terrestrial carbon cycle in the Arctic, due to the large and potentially labile carbon stocks found in permafrost-affected environments, which can lead to a positive climate feedback, along with the possibility of future carbon sinks from northward expansion of vegetation under climate warming. Here we evaluate the simulation of tundra carbon stocks and fluxes in three land surface schemes that each form part of major Earth System Models (JSBACH, Germany; JULES, UK and ORCHIDEE, France). We use a site-level approach where comprehensive, high-frequency datasets allow us to disentangle the importance of different processes. The models have improved physical permafrost processes and there is a reasonable correspondence between the simulated and measured physical variables, including soil temperature, soil moisture and snow.

We show that if the models simulate the correct leaf area index (LAI), the standard C3 photosynthesis schemes produce the correct order of magnitude of carbon fluxes. Therefore, simulating the correct LAI is one of the first priorities. LAI depends quite strongly on climatic variables alone, as we see by the fact that the dynamic vegetation model can simulate most of the differences in LAI between sites, based almost entirely on climate inputs. However, we also identify an influence from nutrient limitation as the LAI becomes too large at some of the more nutrient-limited sites. We conclude that including moss as well as vascular plants is of primary importance to the carbon budget, as moss contributes a large fraction to the seasonal CO_2 flux in nutrient-limited conditions. Moss photosynthetic activity can be strongly influenced by the moisture content of moss, and the carbon uptake can be significantly different from vascular plants with similar LAI.

The soil carbon stocks depend strongly on the rate of input of carbon from the vegetation to the soil, and our analysis suggests that an improved simulation of photosynthesis would also lead to an improved simulation of soil carbon stocks. However, the stocks are also influenced by soil carbon burial (e.g. through cryoturbation) and the rate of heterotrophic respiration, which depends on the soil physical state. More detailed below-ground measurements are needed to fully evaluate soil biological and physical processes. Furthermore, even if these processes are well modelled, the soil carbon profiles cannot resemble peat layers as peat accumulation processes are not represented in the models.

Thus we identify three priority areas for model development: 1. Dynamic vegetation including a. climate and b. nutrient limitation effects. 2. Adding moss as a plant functional type. 3. Improved vertical profile of soil carbon including peat processes.

1 Introduction

Land areas in northern high latitudes may represent a net source or a net sink of carbon to the atmosphere in the future, and there is not yet a consensus as to which of the two is more likely, e.g.

(Cahoon et al., 2012; Hayes et al., 2011). This is not because it is likely to be small: on a pan-Arctic scale we could see anything between a net emission of over 100GtC or a net sink of up to 60GtC by the end of this century (Schuur et al., 2015; Qian et al., 2010). To put this into context, the remaining emissions budget in order to stabilise climate warming below 2°C above pre-industrial levels is less than 250GtC from 2017 (Peters et al., 2015), so it is very important to reduce uncertainty in the northern high latitude carbon cycle. The uncertainty comes largely from the representation of these processes in Earth System Models (ESM's), which are our main tool for future climate projections.

The potential for large carbon emissions comes from the large quantities of old carbon that are frozen into permafrost, protected from decomposition under the current cold climate. Around 800Gt of carbon is stored in permanently frozen soils (Hugelius et al., 2014). If the permafrost thaws, this carbon may decompose and be released to the atmosphere (Burke et al., 2012, 2013; Koven et al., 2015; Schneider von Deimling et al., 2012, 2015; MacDougall and Knutti, 2016). On the other hand, the increased vegetation growth that is already taking place in the Arctic under climate warming (Tucker et al., 2001; Tape et al., 2006) could result in a net uptake of carbon from the atmosphere (Quegan et al., 2011; Qian et al., 2010). It should be noted, however, that in some areas Arctic vegetation growth is not increasing but rather 'browning' (Epstein et al., 2016).

The representations of both permafrost carbon and Arctic vegetation in Earth System Models are not well developed. Some models now include a vertical representation of soil carbon which allows the frozen carbon in permafrost to be included (Koven et al., 2009, 2013; Schaphoff et al., 2013; Burke et al., 2017), but most do not yet represent important mechanisms of carbon storage and release, such as sedimentation, thermokarst formation, and a proper representation of cryoturbation (Schneider von Deimling et al., 2015; Beer, 2016), although sedimentation is included in Zhu et al. (2016). There is also a growing consensus that the chemical decomposition models used in ESMs are not adequate to represent microbial processes (Wieder et al., 2013; Xenakis and Williams, 2014). Vegetation models also, for the most part, do not include the appropriate high latitude vegetation types and those models that have dynamic vegetation are lacking in processes that are essential determinants of vegetation dynamics, such as nutrient limitation and interactions with soil (Wieder et al., 2015).

In this paper we assess the ability of the land surface components from three Earth System Models to represent the observed carbon stocks and fluxes at tundra sites, identifying the processes that have the greatest impact on the uncertainty. These processes are therefore priorities for future model development. Observational studies in tundra environments have shown that carbon dynamics are sensitive to physical conditions (Lund et al., 2012; Cannone et al., 2016; Pirk et al., 2017) , so we first assess the ability of the models to capture the mean physical state of the system and the differences between sites, specifically in terms of snow depth, soil temperature, soil moisture and active layer depth. Secondly, soil carbon stocks are evaluated against measured soil carbon profiles, assessing the main causes of biases in the models. Half-hourly NEE data from eddy flux towers are used to evaluate the simulated carbon fluxes, comparing the models directly against observations before analysing the relationships between ecosystem carbon fluxes and different driving variables. We also consider the impacts of other controlling factors such as nutrient limitation and mosses, whose importance has been identified in previous studies (Atkin, 1996; Uchida et al., 2009) .

This is a synthesis from the recently concluded EU project PAGE21 (Permafrost in the Arctic and Global Effects in the 21st century), evaluating the models that took part in the project (described in Section 2.2, below) at the five PAGE21 primary sites, which are all located in Arctic permafrost regions, specifically Siberia, Sweden, Svalbard and Greenland. After the site-level evaluation of physical processes by Ekici et al. (2015), this evaluation of carbon cycle processes continues site-level model evaluation efforts. The sites are described in detail in Section 2.1.

2 ~~Model descriptions~~Methods

~~The three models studied here are JSBACH, JULES and ORCHIDEE. These are all land surface components of major Earth System Models.They can be run in a coupled mode within the ESM, or, as here, they can be run standalone forced by observed meteorology. Each model had some development of high latitude processes during the PAGE21 project, and model developments have also been ongoing since the conclusion of the project in late 2015 (see below).~~ This study takes three different angles: 1) Comparison with observed indicators. 2) Comparison of processes between models. 3) Comparison of geographical conditions (e.g. vegetation, permafrost) between sites. The structure of the methods section follows this, describing firstly the observational indicators used (Section 2.1), secondly the processes represented in the models (Section 2.2), and thirdly the conditions at the sites (Section 2.1). Lastly, details of the simulation set-up and forcing data are given in Section 2.1.

2.1 ~~JSBACH~~Evaluation data

~~The Jena Scheme for Biosphere-Atmosphere Coupling in Hamburg (JSBACH 3.0 (Raddatz et al., 2007; Brovkin et al., 2009)) is the land surface component of the Max Planck Institute Earth system model (MPI-ESM). The model simulates water fluxes, heat fluxes, and carbon fluxes from vegetation and soil via one-dimensional vertical fluxes. Photosynthesis in JSBACH is based on the approaches of Farquhar et al. (1980) and Collatz et al. (1992) , as described in Knorr (2000) . The carbon cycle is represented by three vegetation pools (active, reserves, wood)and five soil carbon pools which are defined by solubility (Goll et al., 2015) . However, the soil carbon model does not have a vertical dimension.~~

2.1.1 Carbon dioxide flux

~~Hydrological fluxes are simulated by a five-layer scheme (Hagemann and Stacke, 2015) . The model is run as a gridded set of points for large scale simulations. Each grid cell is subdivided into tiles~~

~~which represent different vegetation types and which can vary in fractional cover. During~~ Eddy covariance half hourly CO_2 flux data and related meteorological variables used in this study are archived in the PAGE21 ~~, soil freezing, dynamic snow layers and a simple organic layer were added in JSBACH (Ekici et al., 2014) . In the version used in this paper, the simple organic layer is switched off and replaced by a moss layer with dynamic soil moisture contents and thermal properties (Porada et al., 2016) , and additional soil layers were added in order to represent a 50 m depth . The moss carbon fluxes (photosynthesis, respiration) are also simulated, as in the model described by Porada et al. (2013) . In the version used here, the moss carbon fluxes are not yet fully coupled into the JSBACH carbon cycle, so the moss carbon fluxes are considered separately in the analysis that follows.~~

2.2 ~~JULES~~

~~JULES is the land surface component of the new community Earth System model, UKESM (Jones and Sellar, 2015) . It can also be run offline forced by observed meteorology, and it can be run at a regional or point scale as well as globally. JULES is described in Best et al. (2011); Clark et al. (2011) . It is a community model with many users and many ongoing developments. JULES includes a dynamic vegetation model (TRIFFID), surface energy balance, a dynamic snowpack model (vertical processes only), vertical heat and water fluxes, soil freezing, large scale hydrology, and carbon fluxes and storage in both vegetation and soil . It also includes specific representations of crops, urban heat and water dynamics, fire diagnostics and river routing.~~

~~During PAGE21 the permafrost physics in JULES was improved (Chadburn et al., 2015a) , and a vertical representation of soil carbon, including cryotubation mixing, was added (Burke et al., 2017) . In this work the vertical soil carbon, organic soil properties, deep soil column (including bedrock)and high resolution soil are used. We also use the 9 PFT's described in Harper et al. (2016) and the latest set of PFT parameters from the UKESM project. For more details of soil and vegetation configuration see Simulation Set-up (Section 2.1)and Appendix. The version of JULES used is available on https:~~ fluxes database (http://~~code. metoffice.gov.uk~~www.europe-fluxdata.eu/~~svn~~page21) which is part of the European Flux Database Cluster.

Flux post-processing was performed consistently for all the sites following the protocol applied for the Fluxnet 2015 data release (http:/~~jules~~/~~main~~fluxnet.fluxdata.org/~~branches~~data/~~dev~~fluxnet2015-dataset), with customized choices of the processing options. The applied scheme included: (i) a quality assessment/~~eleanorburke/vn4.3~~quality control procedure over single variables aimed at detecting implausible values or incorrect time stamps (e.g. by comparing patterns of potential and observed downward shortwave radiation at a given location); (ii) the computation of net ecosystem exchange (NEE) by adding the CO_2 flux storage term calculated from a single CO_2 concentration measurement point (at the top of the flux tower) and assuming a vertically uniform concentration field; (iii) the de-spiking of NEE based on Papale et al. (2006) using a threshold value (z=5); (iv) NEE filtering according to an ensemble of friction velocity (u*) thresholds obtained by bootstrapping following the methods of Barr et al. (2013) and Papale et al. (2006) and selection of a u* threshold, different for each year, based on the highest model efficiency (Nash-Sutcliffe); (vi) the gap-filling of NEE time series with the marginal distribution sampling (MDS) method (Reichstein et al., 2005) .

Finally, NEE was partitioned into the gross primary productivity (GPP) and ecosystem respiration (Reco) components using a semi-empirical model based on hyperbolic light response curve fitted to daytime NEE data (Lasslop et al., 2010) . The years of data available for each site are given in supplementary Table S1.

2.1.1 Soil carbon profiles

Typical soil profiles with data on soil organic carbon content were generated for each site. Based on extensive field campaigns in each study area, individual pedons for representative landscape and soil types were combined and harmonized. In brief, soils were classified and sampled from open soil pits dug down to the permafrost. Permafrost samples were collected through manual coring into the permafrost at the bottom of the soil pit. In most cases, soils were sampled to a depth of 1 m. The harmonized soil profiles were generated by averaging several soil pedons per landscape type at a 1 cm depth resolution. For more detailed descriptions of field sampling and laboratory procedures see Palmtag et al. (2015); Siewert et al. (2015, 2016) . Top 1m total soil carbon values were calculated from a weighted average of different typical profiles, based on the fractional coverage of landscape types in the footprint area of the flux towers.

2.2 ~~ORCHIDEE~~

2.1.1 Snow depth

Snow depth was recorded using automatic sensors (except Abisko where it is manual). Snow depth from the Abisko mire (Storflaket) was recorded manually monthly (Johansson et al., 2013) . Snow depth at Samoylov and Bayelva was recorded hourly, and for Zackenberg 3-hourly (using sonic range and laser sensors). Snow depth at Kytayk was measured by means of a 70 cm vertical profile made of thermistors spaced every 5 cm (2.5 cm between 0 and 10 cm height from the ground). Data were logged every 2 hours and the snow-air interface level was identified by analyzing the profile patterns with a Matlab® routine calibrated to search for deviations between consecutive resistance readings above a given threshold. Years used for each site are given in supplementary Table S1.

~~ORCHIDEE is the land-surface component of the IPSL climate model as well as a standalone land surface model. ORCHIDEE simulates the principal processes of the biosphere influencing the global carbon cycle (photosynthesis, autotrophic and heterotrophic respiration of plants and in soils, fire, etc.) as well as latent, sensible, and kinetic energy exchanges at the land surface (Krinner et al., 2005) .~~

2.1.2 Soil temperature

For Samoylov, Bayelva, Kytalyk and Zackenberg, soil temperature was recorded hourly using thermistors (Kytalyk set-up described in van der Molen et al. (2007)). Ground temperatures for Abisko mire were recorded at the Storflaket mire, at boreholes cased with plastic tubes and instrumented with Hobo loggers U12 (Industry, 4 channels) together with Hobo soil temperature sensors (Johansson et al., 2011) . Years used for each site are given in supplementary Table S1.

~~The ORCHIDEE high-latitude version includes vertically resolved soilcarbon and cryoturbative mixing (Koven et al., 2009) , a scheme describing soil freezing and its effect on soil~~

2.1.3 Soil moisture

Continuous soil moisture measurements are only available for Bayelva, Samoylov and Zackenberg. At Samoylov and Bayelva, hourly volumetric soil water content was recorded (using Time Domain Reflectometry). At Zackenberg soil moisture was measured using permanently installed ML2x Thetaprobes (Lund et al., 2014) . Years used for each site are given in supplementary Table S1. Indicative soil moisture levels for Abisko mire were collected from May to October 2015 (Pedersen et al., 2017) , measured manually as volumetric soil water content integrated over 0-6 cm depth using a handheld ML2x Theta Probe (Delta-T Devices Ltd., Cambridge, UK). Soil moisture was measured 5 times in each plot and averages were subsequently used.

2.1.4 Active layer depth

Active layer depth was measured at CALM grids at most of the sites. At Bayelva there is no CALM grid, so active layer was estimated from soil temperature measurements and is given as an 'indicative' value. Active layer thickness monitoring is determined by mechanical probing. A 1 cm diameter graduated steel rod is inserted into the soil to the depth of resistance to determine the active layer thickness (Åkerman and Johansson, 2008) according to the CALM standard.

2.1.5 Leaf area index

Leaf area index was taken from MODIS product (MODIS15A2) , for the closest coordinates to the sites. This product has been successfully applied to tundra sites (Cristóbal et al., 2017) . It was evaluated by Cohen et al. (2006) who found an RMSE of 0.28 at a tundra site. There are, however, still considerable uncertainties in using this data product (see Section 3.6.1).

2.1.6 GPP per unit leaf area

This was calculated using the partitioned GPP from the eddy covariance data (Section 2.1.1), averaged daily and taken on the same day as the values from the MODIS LAI product (Section 2.1.5). Note that there are no time-resolved GPP values for Bayelva due to insufficient data. The extracted GPP values were divided by the appropriate LAI estimates and the resulting values were collected for all sites and binned into intervals of air temperature (1.5°C) and shortwave radiation (20 Wm^{-2}), for which the mean and standard deviation were then calculated (shown on Figure 9).

2.2 Model description

The three models studied here are JSBACH (Jena Scheme for Biosphere-Atmosphere Coupling in Hamburg, Raddatz et al. (2007); Brovkin et al. (2009)), JULES (Joint UK Land Environment Simulator, Best et al. (2011); Clark et al. (2011)) and ORCHIDEE (ORganizing Carbon and Hydrology In Dynamic Ecosystems Environment, Krinner et al. (2005)). These are all land surface components of major ESM's (JSBACH: MPI-ESM; JULES: UKESM; ORCHIDEE: IPSL).

These models can be run in a coupled mode within the ESM, or, as here, they can be run standalone forced by observed meteorology. The models are run as a gridded set of points for large scale simulations, and they can also be run for single points, as in this study. Each model had some development of high latitude processes during the PAGE21 project, and model developments have also been ongoing since the conclusion of the project in late 2015.

All the models simulate vertical fluxes of water, heat and carbon between the atmosphere, the vegetation and the soil. Of relevance to permafrost physics, the models simulate a dynamic snowpack by means of a multilayer snow scheme, and the freezing and thawing of soil (Ekici et al., 2014; Gouttevin et al., 2012a; Wang et al., 2013; Best et al., 2011) . All models use a vertical discretisation of soil thermal and hydrological ~~dynamics (Gouttevin et al., 2012a) , and a multi-layer snow scheme with improved representation of snow thermal conductivity, as well as snow settling, water percolation and refreezing (Wang et al., 2013) . In its latest version used in this study, the impacts of soil organic matter on soil thermal and hydraulic properties , including porosity, thermal conductivity, heat capacity and water holding capacity, are incorporated in the model, generally following Lawrence and Slater (2008) . The~~ fluxes, with differing resolutions (see Appendix Table A.2). JSBACH has the lowest resolution soil, with only 5 layers in the top 10 m (Hagemann and Stacke, 2015) , although in this latest version it is extended to 50 m depth with addtional layers. ORCHIDEE and JULES also simulate an extra thermal-only column on the base of the hydrological column, to represent bedrock (Chadburn et al., 2015a) .

Soil thermal and hydrological properties in both JULES and ORCHIDEE have been adapted to allow better representation of organic soils, whereas in JSBACH only mineral soil properties are represented. However, JSBACH additionally simulates a moss/lichen layer at the surface with dynamic moisture contexts and thermal properties (Porada et al., 2016) , which physically represents the surface organic layer. Organic soil properties in JULES are described in (Chadburn et al., 2015a) . In ORCHIDEE the scheme follows Lawrence and Slater (2008) , using the observation-based soil ~~organic~~ carbon map from ~~NCSCD (Hugelius et al., 2014) is used in the thermal and hydrological modules to derive the above mentioned soil properties, after linear interpolation from their original~~

~~4-layer (i.e. 0-30, 30-100, 100-200, 200-300 cm) values to fit ORCHIDEE vertical layers. The latest ORCHIDEE now has the same vertical discretization scheme for the thermal and hydrological modules above 2 m (11 layers), while the thermal module further extends to 38 m (total 32 layers)~~ Hugelius et al. (2014) .

Soil carbon is represented by a multi-pool scheme in all the models, with inputs from vegetation, and decomposition rates depending on soil temperature, soil moisture, and intrinsic turnover times of different pools (Goll et al., 2015; Clark et al., 2011) . Both ORCHIDEE and JULES represent a vertical profile of soil carbon (discretised in line with the soil hydrology), including cryoturbation mixing (Koven et al., 2009; Burke et al., 2017) . JSBACH, on the other hand, represents only a single layer, with decomposition rates determined by conditions in the upper layer of soil.

None of the models simulate nitrogen or other nutrients. Vegetation growth and productivity is therefore only determined by soil moisture and atmospheric forcing data, with no nutrient limitation. Different land cover types are represented in the models by surface tiles, which can vary in fractional cover. In JULES, a dynamic vegetation model is run with 9 competing plant functional types (PFT's) (Harper et al., 2016) , whereas in the other models, vegetation is fixed, but with dynamic phenology. ORCHIDEE has 13 PFT's ~~,~~ but there is no specific high-latitude PFT in the version used here, so C3 grasses are prescribed ~~as a fixed land cover (but with dynamic phenology)~~ for these sites. In JSBACH there are 20 PFT's (including crop and pasture) and for these sites a 'tundra' PFT is used, which is similar to C3 grass but with reduced Vcmax. In JSBACH there is also a dynamic moss model simulating moss photosynthesis and respiration, as in the model described by Porada et al. (2013) . This model represents both mosses and lichens by one plant functional type with 'average' physiological properties. In the version used here, the moss carbon fluxes are not yet fully coupled into the JSBACH carbon cycle, so the moss carbon fluxes are considered separately in the analysis that follows.

For more details of soil and vegetation configuration see Simulation Set-up (Section 2.1) and Appendix.

3 ~~Site descriptions~~

2.1 Site descriptions

The sites represent a range of climatological and biogeophysical conditions across the tundra. Abisko is the warmest site, ~~in the sporadic permafrostzone~~with sporadic permafrost, followed by Bayelva, which is a high Arctic maritime site (on Svalbard), and Zackenberg, which is a maritime site in Greenland (colder than Bayelva). Samoylov and Kytalyk have a continental Siberian climate and the coldest mean annual temperatures. ~~The soil types, vegetation types and the wetness of the ground all vary between sites. The landscapes at each site also differ~~ The landscapes differ between sites, which can influence the permafrost and carbon dynamics, for example ~~via wind-blown snow and lateral~~

~~water fluxes~~through the impact of topography on snow distribution and hydrology. The following sections provide a short description of each study area, and the important climatic and permafrost variables are given in Table 2.

2.2 ~~Abisko~~

At all sites there has been some tendency towards air temperature warming, which in many cases is accompanied by warming or thawing of permafrost (Callaghan et al. 2010; Christiansen et al. 2010 Parmentier et al. 2011; Boike et al. 2013; Lund et al. 2014; Abermann et al. 2017).

2.1.1 Abisko

The Abisko site ~~(68°21' N, 18°49' E, 385m a.s.l) is located about 200 km north of the Arctic Circle~~ is located in the Torneträsk catchment, northernmost Sweden. ~~The catchment ranges from 345 m a.s.l. to 1700 m a.s.l. and is centered around Lake Tornetrsk. Mean annual air temperature is close to 0°C (-0.6°C for the period 1913-2006), and warming has resulted in mean annual air temperatures above 0°C for the last decade (Callaghan et al. (2010) ; Abisko Station meteorological data; www.polar.se/abisko). The Abisko area is situated in a rain shadow and the total annual precipitation was 304 mm for the period 1961-1990 (Alexandersson et al., 1991) . However the total annual precipitation has increased since then and is now around 350 mm (Abisko Station meteorological data; www.polar.se/abisko).~~

~~The vegetation cover in the Abisko area ranges from remnants of boreal pine forest, through the subalpine zone dominated by mountain birch forest, through the low alpine belt, which extends from the treeline up to where *Vaccinium myrtillus* no longer persist, to the high alpine belt with non-vegetated surfaces (Carlsson et al., 1999; Lantmäteriet, 1997) . The footprint of the eddy covariance tower is charaterized by wet fen with no permafrost present, and vegetation dominated by tall graminoids (Jammet et al., 2015, 2017) .~~

According to Brown et al. (1998), the Abisko area lies within the zone of discontinuous permafrost. However, with the observed permafrost degradation during the last decades (Åkerman and Johansson, 2008; Johansson et al., 2011), the area is now more characteristic of the ~~"sporadic permafrost~~"sporadic permafrost" zone. Permafrost is widespread in the mountains (Ridefelt et al., 2008), but at lower elevations permafrost is only found in peat mires (Johansson et al., 2006).

Data from three sites from the Torneträsk catchment (within an area of 10 km) have been used for this study. The principal sites are Storflaket and Stordalen peat mires. The active layer measurements and the ground temperatures are monitored at the Storflaket site (Åkerman and Johansson, 2008; Johansson et al., 2011) and the carbon monitoring, including the eddy covariance measurement, is carried out at the Stordalen site. These two mire sites are very similar in terms of climate, soil profile and permafrost characteristics. The footprint of the eddy covariance tower is charaterized by wet fen with no permafrost present, and vegetation dominated by tall graminoids (Jammet et al., 2015, 2017) .

For comparison, additional soil temperature data is included from a mineral soil site at the Abisko Scientific Research Station, which is not underlain by permafrost.

2.1.2 Bayelva (Svalbard)

2.2 ~~Bayelva (Svalbard)~~

The study site is located in the high Arctic Bayelva River catchment area, close to Ny-Ålesund on Spitsbergen Island in the Svalbard archipelago. ~~The catchment area lies between two mountains, with the glacial Bayelva River originating from the Brøggerbreen glacier. The West Spitsbergen Ocean Current warms this area to an average air temperature of about −13°C in January and +5°C in July; it also provides about 400 mm of precipitation annually, which falls mostly as snow. The area has experienced a significant warming since the 1960s related to atmospheric circulation patterns and in later years the lack of sea ice during winter (Hanssen-Bauer and Førland, 1998; Førland et al., 2012) .~~ The area is characterized by maritime continuous permafrost. In bioclimatic terms the area represents a semi-desert ecosystem (Uchida et al., 2009).

~~The study site is located on Leirhaugen hill (25 m a.s.l.), on permafrost patterned ground mainly consisting of non-sorted soil circles or mud boils. The ground is mostly bedrock but is partly covered by a mixture of sediments, comprising glacial till and finer glacio-fluvial sediments and clays. The mud boils have bare soil centers (about 1m diameter) and a surrounding rim of vegetation including~~ Vegetation includes low vascular plants (mainly grass, sedge, catchfly, saxifrage and willow), mosses and lichens (Ohtsuka et al., 2006; Uchida et al., 2006). The ~~soils are mineral (described as 'silty loam') with low organic content, although there can be locally high concentrations of organic carbon, for example at the base of the soil profile (Boike et al., 2008a) .~~

~~The area is characterized by maritime continuous permafrost with temperatures around -2 to -3°C. The active layer thickness in general exceeds 1m and can reach as deep as 2m in some areas (Westermann et al., 2010) . Recent recent climatic warming has become manifest in the permafrost temperatures (Christiansen et al., 2010) .~~

~~The~~ ground is mostly bedrock but is partly covered by a mixture of sediments. The study site is located on permafrost patterned ground mainly consisting of non-sorted soil circles or mud boils, with around 60% vegetation cover. The eddy covariance measurements were conducted on Leirhaugen hill~~(78°55.0'N, 11°57.0"E). Additional~~ , and additional meteorological observations and ground temperature measurements are continuously conducted at the Bayelva soil and climate monitoring station (Boike et al., 2003, 2008a; Roth and Boike, 2001) 100m away. Over the past decade the Bayelva catchment has been the focus of intensive investigations on soil and permafrost conditions (Roth and Boike, 2001; Boike et al., 2008a; Westermann et al., 2010, 2011), and the surface energy balance (Boike et al., 2003; Westermann et al., 2009). Details of the measurements are provided in Westermann et al. (2009); Lüers et al. (2014).

2.2 ~~Kytalyk~~

2.1.1 Kytalyk

The Kytalyk site ~~(70°50' N, 147°30' E, 10 m a.s.l.)~~ is located in the Kytalyk reserve, 28 km northwest of the village of Chokurdakh in the Republic of Sakha (Yakutia), Russian Federation. The site is located between the East Siberian Sea ~~(150 km to the North)~~ and the transition zone between taiga and tundra. ~~Based on the data from Chokurdakh airport, the monthly mean air temperatures range between -34.2 °C (January) and +10.4 °C (July). There is a current tendency to warming in particular in autumn (Parmentier et al., 2011) . Annual mean precipitation amounts to 232 mm, of which about half falls as snow.~~

~~Three major topographic levels occur around the measurement site. The highest level in the~~ The area is underlain by ~~'Ice complex deposits' or 'Yedoma': ice-rich silt deposits (Schirrmeister et al., 2002; Gavrilov et al., 2003; Zimo~~ permafrost. The measurement site is located on the bottom of a drained former thermokarst lake, and the site is bordered by the edge of the present river floodplain. Both on the floodplain and the lake bottom a network of ice wedge polygons occurs, in general of the low-centered type. ~~The ice wedge polygons on the lake bottom have broad ridges that may coalesce into low palsa-like plateaus. In between these plateaus a network of diffuse, strongly vegetated drainage channels have developed., This network of plateaus and drainage channels locally masks the original polygon structure. The~~ These form a mosaic of low plateaus and ridges ~~is~~ dominated by *Betula nana*, ~~the~~ and diffuse drainage channels ~~are~~ covered with a meadow-like vegetation of *Eriophorum angustifolium* and *Carex* sp. ~~,~~ There is also hummocky *Sphagnum* with low *Salix* dwarf shrubs, polygon ponds ~~are~~ covered with mosses and *Comarum palustre*, deeper ponds where ice wedges have thawed, and drier areas ~~are~~ covered with *Eriophorum vaginatum* tussocks. The soils generally have a 10-40 cm organic top layer overlying silt. ~~In case of wet sites, the organic layer consists of loose peaty material, composed either of sedge roots or *Sphagnum* peat, depending on the vegetation. Drier sites tend to have a thinner, more compact organic layer.~~

~~The area is underlain by continuous permafrost. The active layer ranges from ~25 cm in dry, peat-covered locations to ~50 cm in wet locations. On the floodplain the active layer may be locally thicker.~~

~~The~~ The eddy covariance tower is located at a distance of ca. 200 m from the research station buildings (van der Molen et al., 2007). The tower footprint covers a wet northwestern and southeastern sector dominated by *Sphagnum* and ponds, while the northeastern and southwestern sectors have drier vegetation types.

2.2 ~~Samoylov~~

2.1.1 Samoylov

~~The Lena River Delta in northern Yakutia is one of the largest deltas in the Arctic. Samoylov Island (72°22'N, 126°28'E)~~ Samoylov Island lies within one of the main river channels in the southern part of the ~~delta and is relatively young, with an age of between 4 and 2 ka BP (Schwamborn et al., 2002) . The annual mean air temperature on Samoylov Island from 1998–2011 was −12.5°C, with the coldest monthly temperatures (January and February) around −30°C, and maximum monthly temperature around 10°C (July and August) (Boike et al., 2013) . The~~ Lena river delta, northern Yakutia. The landscape on Samoylov Island, and in the delta as a whole, has generally been shaped by water through erosion and sedimentation (Fedorova et al., 2015), and by thermokarst processes (Morgenstern et al., 2013). ~~The proportion of the total land surface of the delta covered by surfacewater can amount to more than 25% (Muster et al., 2012) .~~

Continuous cold permafrost underlies the study area to between about 400 and 600 m below the surface. The terrace where the study site is situated is covered in low-centred ice wedge polygons~~.~~ ~~In the depressed polygon centres, drainage is impeded due to the underlying permafrost, leading to~~ , with water-saturated soils or small ponds ~~.~~ in the polygon centres. The mineral soil is generally sandy loam, underlain by silty river deposits, with a $\sim$30cm thick organic layer ~~at the surface (Boike et al., 2013) . The vegetation~~ (Boike et al., 2013) . Vegetation in the polygon centres and at the edge of ponds is dominated by sedges and mosses, and at the polygon rims, various mesophytic dwarf shrubs, forbs and mosses dominate (Kutzbach et al., 2007). ~~The maximum summer leaf coverage of the vascular plants was estimated to be about 0.3, and the leaf coverage of mosses was estimated to be about 0.95 (Kutzbach et al., 2007) .~~ It is estimated that moss contributes around 40% to the total photosynthesis ~~(Kutzbach et al., 2007) .~~

~~Continuous cold permafrost (with a mean annual temperature of -10°C at 10 m depth) underlies the study area to between about 400 and 600 m below the surface. The active layer depth is generally less than 1m, and typical snow depth around 0.2-0.4 m (Boike et al., 2013) . Since observations started in 2006, the permafrost at 10.7 m depth has warmed by > 1.5°C (Boike et al. (2013) ; http://gtnpdatabase.org/boreholes/view/5~~

~~Additional detailed~~ (Kutzbach et al., 2007) . Detailed information concerning the climate, permafrost, land cover, vegetation, and soil characteristics of ~~these islands in the Lena River Delta~~ Samoylov Island can be found in Boike et al. (2013) and Morgenstern et al. (2013). Analysis of the energy balance for the site is found in (Boike et al., 2008b; Langer et al., 2011a, b).

2.2 ~~Zackenberg~~

2.1.1 Zackenberg

The Zackenberg study site is located near the Zackenberg Research Station within the Northeast Greenland National Park~~(74°28'N; 20°33'W). High mountains (> 1000 m a.s.l.)~~ , within the continuous permafrost zone. High mountains surround the Zackenberg valley to the west, east and north, ~~while~~

~~in the south a fjord forms its boundary. The area has been covered by the Greenland Ice sheet several times. The climate is high Arctic with an annual mean air temperature of -9.0°C (1996-2014) and only June, July, August and September have mean monthly temperatures above 0°C. The annual mean temperature has increased by 0.06°C per year since 1996 with most rapid warming occurring during summer months (Abermann et al., 2017) . The mean annual precipitation is 211 mm (1996-2014) of which most falls as snow; the water~~ with a fjord to the south, and snow cover is characterized by large interannual variability (Pedersen et al., 2016) . Water availability is thus regulated by topography and snow distribution patterns. ~~The seasonal snow cover is characterized by large interannual variability with maximum snow depths ranging from 0.13 m in 2013 to 1.33 m in 2002 (Pedersen et al., 2016) .~~

~~Most vegetated surfaces~~ Most vegetation in the Zackenberg valley ~~are~~ is located below 300 m.a.s.l., where the lowland is dominated by non-calcareous sandy fluvial sediments (Elberling et al., 2008)~~.~~ ~~Mineral soil types dominate while~~ , and peat soils have limited spatial coverage (Palmtag et al., 2015). ~~At least five main plant community types can be identified: fens occurring in water-saturated areas (Dupontia psilosantha, Eriophorum scheuchzeri), grasslands in semi-sloping, wet-to-moist terrain (Arctagrostis latifolia, Eriophorum triste), Salix arctica snow-beds mostly in slopes with prolonged snow cover, *Cassiope tetragona* heaths in drier, level ground in the central valley, and Dryas heath in dry and wind-exposed areas (Elberling et al., 2008) .~~ The study site is located within a ~~*C.*~~ *Cassiope tetragona* tundra heath, dominated by *C. tetragona*, *Dryas integrifolia* and *Vaccinium uliginosum*, ~~accompanied by~~ with patches of mosses.

~~Zackenberg is situated within the continuous permafrost zone, and the landscape development is dominated by periglacial processes. Only the upper 45-80 cm of the soil (active layer thickness) thaws every summer. However, in a CALM (Circumpolar Active Layer Monitoring Network) field close to the study site, the maximum thaw depth has increased with 1.0-1.5 cm per year since 1997 (Lund et al., 2014) .~~

Several studies on soil and permafrost (Palmtag et al., 2015; Westermann et al., 2015), surface energy balance (Lund et al., 2014; Stiegler et al., 2016; Lund et al., 2017) and carbon exchange (Mastepanov et al., 2008; Lund et al., 2012; Elberling et al., 2013) have been published based on data from this site. A rich data set is available from this site through the extensive, cross-disciplinary Greenland Ecosystem Monitoring (GEM) programme (www.g-e-m.dk).

3 ~~Methods~~

2.1 ~~Evaluation data~~

2.0.1 ~~Carbon dioxide flux~~

~~Eddy covariance half hourly CO_2 flux data and related meteorological variables used in this study are archived in the PAGE21 fluxes database (http://www.europe-fluxdata.eu/page21) which is part of the European Flux Database Cluster.~~

~~Flux post-processing was performed consistently for all the sites following the protocol applied for the Fluxnet 2015 data release (http://fluxnet.fluxdata.org/data/fluxnet2015-dataset), with customized choices of the processing options. The applied scheme included: (i) a quality assessment/quality control procedure over single variables aimed at detecting implausible values or incorrect time stamps (e.g. by comparing patterns of potential and observed downward shortwave radiation at a given location); (ii) the computation of net ecosystem exchange (NEE) by adding the CO_2 flux storage term calculated from a single CO_2 concentration measurement point (at the top of the flux tower) and assuming a vertically uniform concentration field; (iii) the de-spiking of NEE based on Papale et al. (2006) using a threshold value (z=5); (iv) NEE filtering according to an ensemble of friction velocity (u*) thresholds obtained by bootstrapping following the methods of Barr et al. (2013) and Papale et al. (2006) and selection of a u* threshold, different for each year, based on the highest model efficiency (Nash-Sutcliffe); (vi) the gap-filling of NEE time series with the marginal distribution sampling (MDS) method (Reichstein et al., 2005) .~~

~~Finally, NEE was partitioned into the gross primary productivity (GPP) and ecosystem respiration (Reco) components using a semi-empirical model based on hyperbolic light response curve fitted to daytime NEE data (Lasslop et al., 2010) . The years of data available for each site are given in supplementary Table S1.~~

2.0.1 ~~Soil carbon profiles~~

~~Typical soil profiles with data on soil organic carbon content were generated for each site. Based on extensive field campaigns in each study area, individual pedons for representative landscape and soil types were combined and harmonized. In brief, soils were classified and sampled from open soil pits dug down to the permafrost. Permafrost samples were collected through manual coring into the permafrost at the bottom of the soil pit. In most cases, soils were sampled to a depth of 1 m. The harmonized soil profiles were generated by averaging several soil pedons per landscape type at a 1 cm depth resolution. For more detailed descriptions of field sampling and laboratory procedures see Palmtag et al. (2015); Siewert et al. (2015, 2016) . Top 1m total soil carbon values were calculated from a weighted average of different typical profiles, based on the fractional coverage of landscape types in the footprint area of the flux towers.~~

2.0.1 ~~Snow depth~~

~~Snow depth was recorded using automatic sensors (except Abisko where it is manual). Snow depth from the Abisko mire (Storflaket) was recorded manually monthly (Johansson et al., 2013) . Snow height at Samoylov and Bayelva was recorded hourly, and for Zackenberg 3-hourly (using sonic~~

~~range and laser sensors). Snow depth at Kytayk was measured by means of a 70 cm vertical profile made of thermistors spaced every 5 cm (2.5 cm between 0 and 10 cm height from the ground). Data were logged every 2 hours and the snow-air interface level was identified by analyzing the profile patterns with a Matlab® routine calibrated to search for deviations between consecutive resistance readings above a given threshold. Years used for each site are given in supplementary Table S1.~~

2.0.1 ~~Soil temperature~~

~~For Samoylov, Bayelva, Kytalyk and Zackenberg, soil temperature was recorded hourly using thermistors (Kytalyk set-up described in van der Molen et al. (2007)). Ground temperatures for Abisko mire were recorded at the Storflaket mire, at boreholes cased with plastic tubes and instrumented with Hobo loggers U12 (Industry, 4 channels) together with Hobo soil temperature sensors (Johansson et al., 2011) . Years used for each site are given in supplementary Table S1.~~

2.0.1 ~~Soil moisture~~

~~Continuous soil moisture measurements are only available for Bayelva, Samoylov and Zackenberg. At Samoylov and Bayelva, hourly volumetric soil water content was recorded (using Time Domain Reflectometry). At Zackenberg soil moisture was measured using permanently installed ML2x Thetaprobes (Lund et al., 2014) . Years used for each site are given in supplementary Table S1. Indicative soil moisture levels for Abisko mire were collected from May to October 2015 (Pedersen et al., 2017) , measured manually as volumetric soil water content integrated over 0-6 cm depth using a handheld ML2x Theta Probe (Delta-T Devices Ltd., Cambridge, UK). Soil moisture was measured 5 times in each plot and averages were subsequently used.~~

2.0.1 ~~Active layer depth~~

~~Active layer depth was measured at CALM grids at most of the sites. At Bayelva there is no CALM grid, so active layer was estimated from soil temperature measurements and is given as an 'indicative' value. Active layer thickness monitoring is determined by mechanical probing. A 1 cm diameter graduated steel rod is inserted into the soil to the depth of resistance to determine the active layer thickness (Åkerman and Johansson, 2008) according to the CALM standard.~~

2.0.1 ~~Leaf area index~~

~~Leaf area index was taken from MODIS product (MODIS15A2) , for the closest coordinates to the sites. This product has been successfully applied to tundra sites (Cristóbal et al., 2017) . It was evaluated by Cohen et al. (2006) who found an RMSE of 0.28 at a tundra site. There are, however, still considerable uncertainties in using this data product (see Section 3.6.1).~~

2.0.1 ~~GPP per unit leaf area~~

~~This was calculated using the partitioned GPP from the eddy covariance data (Section 2.1.1), averaged daily and taken on the same day as the values from the MODIS LAI product (Section 2.1.5). Note that there are no time-resolved GPP values for Bayelva due to insufficient data. The extracted GPP values were divided by the appropriate LAI estimates and the resulting values were collected for all sites and binned into intervals of air temperature (1.5°C) and shortwave radiation (20 Wm^{-2}), for which the mean and standard deviation were then calculated (shown on Figure 9).~~

2.1 Simulation set-up

The sites were represented in all the models by a single vertical column, although there was some horizontal representation by means of tiling approaches (see model ~~descriptions~~description, Section 2.2). The models were run in the most 'up-to-date' configurations, including new permafrost-relevant model developments where available. Variables were output at hourly and/or daily resolutions.

The meteorological driving data were prepared using observations from the site combined with reanalysis data for the grid cell containing the site. For the period 1901-1979, Water and Global Change forcing data (WFD) was used (Weedon et al., 2011). Data is provided at half-degree resolution for the whole globe at 3-hourly time resolution from 1902-2001. For the period 1979-2014, WATCH Forcing Data Era-Interim (WFDEI) was used (Weedon, 2013). For the time periods where observed data were available, correction factors were generated by calculating monthly biases relative to the WFDEI data. These corrections were then applied to the time-series from 1979-2014 of the WFDEI data. The WFD before 1979 was then corrected to match this data and the two datasets were joined at 1979 to provide gap-free 3-hourly forcing from 1901-2014. Local meteorological station observations were used for all variables except snowfall, which was estimated from the observed snow depth by treating increases in snow depth as snowfall events with an assumed snow density (see Appendix). These reconstructions were then used to provide correction factors to WFDEI and WFD. This leads to a more realistic snow depth in the model than using direct precipitation measurements, due to wind effects and the difficulty of accurately measuring snowfall. However, the local precipitation measurements were still used for rainfall, as this is much more reliable, with a potential undercatch of only around 10% (Yang et al., 2005) . For Abisko, meteorological data from the research station were used, but additionally corrected by scaling the snowfall according to the ratio of monthly snow depths at the mire vs the research station (snow depth was only measured monthly at Storflaket mire), and a reduction of 1°C in air temperature. Even with these corrections, there is still considerable uncertainty in precipitation forcing, particularly the snowfall, so in order to test the impact of this, two of the models (JULES and JSBACH) performed two additional sets of simulations, with snowfall increased and reduced by 50%.

Spin-up was performed as consistently as possible between the models, using the meteorological forcing from 1901-1930. Years were selected at random from this 30 year period and the models were run for 10000 years with pre-industrial CO_2 (1850, 286 ppm), followed by 50 years with changing CO_2 (1851-1900). The model state at the end of this spin-up period was taken as the initial state for the main run (1901-01-01 to 2013-12-31). For JSBACH, there was an initial 50 years of hydrological spin-up before the main spin-up, with the permafrost impact on hydrology switched off, to allow the water to form a realistic profile (permafrost layers are impermeable and thus unrealistic initial conditions could otherwise be preserved). For JSBACH, the long spin-up was also between 7000-8000 years rather than 10000, since in this model there is no vertical representation of soil carbon, and therefore the soil carbon pools equilibriate much more quickly and had reached a steady state after 7-8000 years. The CO_2 forcing data is from Meinshausen et al. (2011).

The soil parameters in the models were set up to represent each site as closely as possible (see Appendix, and Table A.1). These drew from literature values, a PAGE21 deliverable 'Catalogue of physical parameters', and field experience. (Note that the soil carbon profiles described in Section 2.1.1 were not used for this).

Vegetation was prescribed in ORCHIDEE and JSBACH. Since these are tundra sites, JSBACH used a 'tundra' PFT (100% coverage), which is similar to C3 grass but with reduced Vcmax (maximum rate of carboxylation in leaves). ORCHIDEE prescribed C3 grass (100% coverage) as there is no tundra PFT in this model version. JULES was run with dynamic vegetation using 9 PFT's (Harper et al., 2016), which do not include any tundra PFT's. All 9 PFT's prognostically determine their coverage according to the environmental conditions, and they are all allowed to compete for space. In practice, only the C3 grass PFT is able to grow at these sites.

Some experiments were performed to separate the impacts of different processes. ORCHIDEE was run with and without vertical mixing of soil carbon. JSBACH carbon fluxes were analysed with and without an additional contribution from a new moss photosynthesis scheme. In JULES, an extra set of simulations was performed with fixed vegetation, to compare with the dynamic vegetation scheme.

3 Results and discussion

The carbon dynamics are intrinsically linked to the physical state of the system ~~,~~ (for example, determining the rate of soil carbon decomposition), so we start by assessing the snowpack, soil temperature, soil moisture, and active layer thickness in all three models. The model physics has also been evaluated in detail in previous publications (Ekici et al., 2015, 2014; Chadburn et al., 2015a; Porada et al., 2016), so is kept short here. In these studies, representing organic soil was identified as a key influence on the simulation of soil physics, and following this we compare organic against mineral soils in our analysis. We then evaluate the soil carbon stocks and the ecosystem CO_2 fluxes, and we analyse the CO_2 fluxes in detail. The fluxes depend on every part of the system, so all of the preceding analysis contributes to our understanding of the carbon dynamics at these sites.

3.1 Snow

~~Seasonal~~ The seasonal cycle of snow depth is shown in Figure 1. It depends strongly on the snowfall driving data. Since the snowfall was back-calculated from the snow depth, the accumulation period should match well with observations. There is still some variation due to the fresh snow density in the models (which can differ both from the assumed density in making the driving data, and between the models), and furthermore the compaction of the snow is dependent on the model process representation and physical conditions. Nonetheless, ~~the models all~~ for the most part the models make a reasonable simulation of the snowpack accumulation and compaction~~. However, during the melting season they~~ , with the exception of Abisko where the models are all biased high. Here, snow inputs are particularly uncertain as no high-resolution timeseries of snow depth are available (unlike the other sites). We performed a sensitivity study to test the impact of uncertainties or variability in snow depth on the simulated carbon-cycle processes. In this study, a reduction of 50% in snowfall allows the models to simulate a realistic snow depth at Abisko – see supplementary material. The impacts on soil carbon stocks and fluxes are fairly small, however (between 0.2% and 10%, supplementary Figures S7 and S8).

During the melting season the models are less accurate than during accumulation, with the snow often melting ~~a little too early~~too early, by up to 25 days in the most extreme case. Our method of back-calculating snowfall from snow depth may miss some snowfall events during the melt season. There are also many other potential influences such as albedo effects, snow-vegetation interactions and the influence of wind-blown sediment. For example, the vegetation in the models is quite tall (up to 1m), and can lead to a lower albedo in the models than reality, and thus faster snowmelt (this is modelled by interpolating between snow-covered and snow-free albedo depending on snow depth and vegetation height). At Bayelva, where the vegetation is particularly small (∼5cm), there is a notable underestimation of the snow depth and early snowmelt in all models, which supports this hypothesis (snow at Bayelva can be modelled very well when vegetation is not included (López-Moreno et al., 2016)). Snowdrift is only represented by scaling the snowfall data to match the observed snow accumulation, which limits the extent to which snowpack dynamics can be recreated by the models.

It is important to be careful when modelling snow depth based on single point observations, as they may not be representative of the area as a whole. Further details on the representativity of snow depths are given in supplementary information. The sensitivity of carbon cycle processes to increased/reduced snowfall is discussed in Sections 3.5 and 3.6.1.

3.2 Soil temperature

Soil temperature annual cycles at ~40cm depth are shown on Figure 2. In general the models simulate the soil temperature at mineral soil sites quite well: Bayelva and Zackenberg sites on Figure 2. There are greater errors in the simulation of organic soils: Abisko, Kytalyk and Samoylov on Figure 2.

For JSBACH and ORCHIDEE, the annual cycles of temperature are too large for the organic sites, indicating that these models need to better represent the insulating/damping properties of organic soils. To illustrate this, additional observations are shown on the Abisko plot (Fig. 2), from mineral soil at the nearby research station (where there is no permafrost). This line matches much more closely with the ORCHIDEE and JSBACH simulations, suggesting that these models are behaving thermally like a mineral soil. At Abisko, permafrost only occurs in peat plateaus and thus including organic soil properties in the models is essential for capturing the difference between permafrost and non-permafrost conditions.

In JULES, on the other hand, the annual cycle amplitude is too small at the organic sites and also at Zackenberg, mostly due to biases in the winter soil temperatures. This suggests that the snow thermal conductivity or density may be too low in JULES. A similar problem was found with a previous JULES simulation of Samoylov island, using a similar model set-up and forcing data (Chadburn et al., 2015a). There, the winter soil temperature was improved by increasing snow density. Indeed, the conductivity of snow in the JULES simulations is between 0.03-0.1 $Wm^{-1}K^{-1}$ at the sites with shallow snow (and in the upper layers of the snowpack at sites with deeper snow), which is considerably lower than typical values for similar tundra sites, which suggest a realistic conductivity would be around 0.2-0.3 $Wm^{-1}K^{-1}$, at least for the upper part of the snowpack (Gouttevin et al., 2012b; Domine et al., 2016) . See supplementary material for further discussion on snow conductivity/density.

3.3 Soil moisture

As with temperature, the (unfrozen) soil moisture is simulated well at mineral soil sites - see Bayelva and Zackenberg in Figure 3. In the winter, ORCHIDEE ~~has a problem in that it~~ does not represent the unfrozen water fraction in frozen soils, but the other models simulate a reasonable water content in winter. However, soil moisture is in general too low at organic sites - Samoylov and Abisko mire. The soils should be able to hold water near the surface and remain saturated very close to the surface (or even above). This points to problems with the hydrology schemes. The soil moisture is very important for the soil temperatures, and it can also have a strong influence on soil carbon stocks and the partitioning of decomposition into CO_2 and methane. Furthermore, it ~~is important for moss photosynthesis, and therefore~~ influences vegetation growth, and thus the uptake of CO_2 from the atmosphere. Therefore it is important to further improve the soil hydrology in these models.

Note that saturated zones can be influenced by landscape heterogeneity and lateral water fluxes that would not be captured in a point simulation. This can potentially be simulated by the models as a landscape average (see for example Gedney and Cox (2003)). However, such schemes simulate only a gridbox mean water content, which does not capture, for example, the influence of anaerobic conditions on decomposition.

Figure 3 shows quite a large variation in the timing of freeze-up and thaw between the models, reflecting the soil temperature differences in Figure 2. Correspondingly, the largest differences are at the organic soil sites.

3.4 ALT

The active layer depth is shown on Figure 4. In the models it is calculated by interpolation of soil temperatures to find the daily thaw depth, except in JULES which uses the method of Chadburn et al. (2015a). (The two methods differ at most within the thickness of the soil layers, Table A.2). In ORCHIDEE and JSBACH the active layer is too deep, which corresponds to the too-warm soil temperatures in summer, Fig. 2. In JSBACH the summer temperatures are only a little warmer than the observations - certainly closer than in ORCHIDEE, yet at some sites the active layer is just as deep. This is because technically the ALT cannot be diagnosed correctly in JSBACH, given the thick soil layers below 20 cm depth (see Appendix Table A.2). Increasing the resolution of the soil layers, while it does not make a big difference to the soil temperature profile, has a very large impact on the simulation of the active layer depth, as shown by Chadburn et al. (2015b). In JULES there is generally quite a good match to the observations as supported by the fact that the summer soil temperatures match closely with the observations for most sites. For Zackenberg the active layer is a little too shallow, but still in the range of observed values. This shows the importance both of resolving the soil column and the insulating effects of organic matter for determining the summer soil temperatures (Dyrness, 1982).

3.5 Soil carbon stocks

JULES and ORCHIDEE represent a vertical profile of soil carbon, whereas JSBACH does not. Without a vertical representation of soil carbon it is not possible to simulate permafrost carbon stocks, because all of the carbon is subject to the seasonal freezing and thawing of the active layer and the model does not contain any ‘inert’ permanently frozen carbon. Therefore, a vertical representation of soil carbon is prerequisite for simulating soil carbon stocks at these sites. However, JULES and ORCHIDEE have some problems in simulating the profiles - Figure 5. The ~~biggest~~ most obvious problem is underestimation: there is ~~very~~ much too little carbon simulated at many of the sites (see last panel on Figure 5). For the sites where the quantity of soil carbon is somewhat realistic, the shape of the profiles vary from a steep exponential-looking decay with depth, to a shallower decline with more carbon in the deeper soil. The same kind of profiles are seen in the observations, particularly for the mineral soil sites (Bayelva and Zackenberg). However, neither of the models can produce the carbon-rich peaty layers of the organic soils. To simulate this would require additional process representation in the models, including representing saturated (and thus anaerobic) conditions in peat soil, and a dynamic representation of bulk density.

The reasons for the major underestimation are different in JULES and ORCHIDEE. In JULES, the main problem is that the GPP is underestimated, so there are not enough plant inputs to accumulate carbon in the soil. This is made clearer by Figure 6, which shows the relationship between GPP and top 1m soil carbon stocks ~~.~~In JULES, the relationships are very similar to the observations, which indicates that the turnover of carbon in the soil is reasonable in JULES. Therefore, if the GPP were large enough, the soil carbon stocks would be much more realistic. In ORCHIDEE, the story is different. Even when the vegetation is productive, the soil carbon stocks are still very low. This indicates a problem with the soil carbon decomposition. There are two factors that could affect this. Firstly, the soil temperatures in ORCHIDEE are much too warm, and the active layer is too deep (Fig.s 2 and 4). This can lead to too much decomposition. In order to improve this the model needs to better represent the insulation from the organic soils. Another possible problem is the deep soil respiration. In ORCHIDEE the only factor that suppresses the soil respiration at depth is the cold and/or frozen nature of the ground. In JULES, however, there is an additional decay of respiration with depth that empirically represents some processes that are missing in the model (following the implementation in CLM, see Koven et al. (2013)). Including this in ORCHIDEE could lead to a higher carbon stock at depth. The deeper soil carbon stocks are also influenced by long-term burial processes, which are only represented by a simple diffusion scheme in these models. We include JSBACH on Figure 6 because the top 1m soil carbon is mostly in the active layer. ~~However, given~~ Given that the decomposition in JSBACH is controlled by the temperature of the top soil layer (3cm), it is not surprising that the ~~model somewhat underestimates the carbon stocks.~~ relationships are not captured perfectly, as the upper soil layer will be much more sensitive to variations in temperature than the deeper ones. However, on average the turnover is quite realistic for this model.

It should be noted that the observed relationship on Figure 6 may be confounded by the history of soil carbon formation at these sites. There is inconsistency between Holocene climate and the pre-industrial climate used in model spin-ups. Reconstructed Holocene climate for northern hemisphere is warmer than pre-industrial (Marcott et al., 2013), and possibly wetter, favouring the formation of peat, so some underestimation by the models may be expected.

The soil carbon stocks are sensitive to changes in snow depth in these models (see supplementary Figure S8), through changes in soil temperature (JSBACH) and changes in vegetation growth (JULES). In JULES, both vegetation and soil temperature changes affect the soil carbon, but the vegetation effect dominates. In fact, for two of the sites (Kytalyk and Samoylov), the vegetation coverage is so different during spinup that the simulation with increased snowfall accumulates twice as much soil carbon as the default case (although the stocks are still much too small and the absolute difference is less than 10 kgm^{-2} in the whole soil column).

We conclude that improving soil carbon stocks demands a different priority in each model. For JULES, the first priority is to simulate realistic vegetation productivity, for ORCHIDEE it is to improve the soil carbon decomposition, and for JSBACH it is to represent a vertical profile of soil carbon. Assuming we can combine the best features from all of the models, the greatest difference between the observed and simulated profiles will be the peaty, organic layers that are present in observations and not models (Figure 5). Therefore the next priority for model development is to better represent these organic soils. See e.g. Frolking et al. (2010); Schuldt et al. (2013) for examples of modelling peat. While peatlands represent a small fraction of the land surface, they contain very large carbon stocks (Yu et al., 2010), so it is important to include them in ESM's.

3.6 Carbon fluxes

Figure 7 shows the seasonal cycle of CO_2 flux at every site. The day-time and night-time fluxes are plotted separately (partitioned by incoming shortwave radiation), showing in general uptake during the day and emissions during the night. For the most part the models show uptake and emissions at the same time as the observations, and a similar timing of peak uptake/emission (one exception being the spring daytime flux in ORCHIDEE, see Section 3.6.1).

From the observations we also have the gap-filled estimates of annual gross primary productivity (GPP) and ecosystem respiration (Reco), which are compared with the annual totals for each model on Figure 8 (the moss GPP shown here is discussed in Section 3.6.3). For the GPP we see that for each model there is a positive correlation (sites with larger GPP in reality have larger GPP in the models), but that the overall values are too small for JULES, for ORCHIDEE there is a bigger variation, and for JSBACH, they tend to be too large for the less productive sites and too small for the more productive sites - i.e. the slope of the relationship between model and observations is too shallow. Nonetheless, a significant amount of the variation between sites is captured by the models, to which the only inputs are climate data and soil properties. Of these, climate is the main driver of vegetation growth in these models (since nutrient limitation is not included, the soil only impacts the vegetation through moisture stress - which is also partly climate-related), so we can say that a lot of the difference between the GPP/Reco across different sites is due to the difference in climate. In fact, in JULES and JSBACH, over 90% of the variation in GPP between sites is explained by the model, despite the systematic biases (R squared values of modelled GPP against observed GPP: JSBACH - 0.94, JULES - 0.95, ORCHIDEE - 0.63). This suggests that a model based on climate alone and with one tundra PFT could capture most of the variability in tundra carbon uptake, if the vegetation was correctly calibrated. This is a promising sign that the model simulations could be easily improved.

Due to the magnitude of errors in GPP and Reco, when considering the difference between the two - the net ecosystem exchange (NEE), the noise will be larger than the signal. Nonetheless, the models and observations both generally show a carbon sink in the present day, due to environmental conditions being more favourable for growth (warmer, more CO_2) than in the 'pre-industrial' spin-up period (Table 3).

3.6.1 Drivers of carbon fluxes

The models indicate different drivers of GPP in different parts of the growing season. In particular, ~~that GPP depends mostly on LAI until around the middle of the growing season (end of July) and mostly on shortwave radiation~~ the increase in GPP in the first half of the season is driven by increasing LAI, and the downward trend of GPP in the second half of the season ~~(August onwards)~~is driven by shortwave radiation. There is also a temperature dependence in all parts of the growing season. These relationships are shown in Supplementary Figure S1. Figure S1 also shows the plant respiration in the models, which exhibits a similar behaviour to the GPP, being influenced by temperature, shortwave radiation and LAI. The fact that these variables influence the GPP and autotrophic respiration is clear from the model structure (for example Knorr (2000); Clark et al. (2011)), however the apparent split between the two halves of the season is an emergent behaviour.

The other component of the ecosystem respiration is heterotrophic respiration. This does not exhibit the same dependencies as the plant respiration as it is determined by below-ground conditions. The heterotrophic respiration has a loose relationship with air temperature and a much stronger relationship with the $\sim$20cm soil temperature - see Supplementary Figure S2.

In order to compare the photosynthesis schemes in the models more directly, we normalise by the LAI. It then becomes clear that the photosynthesis models in JSBACH and ORCHIDEE are in fact quite similar. Figure 9 shows the normalised GPP (per m^2 of leaf) against the air temperature and shortwave radiation. JSBACH and ORCHIDEE show similar relationships, although ORCHIDEE still has a slightly higher GPP, potentially explained by the fact that Vcmax is higher. On these plots we also show the limited data that we can plot from observations, using MODIS LAI. It is clear that the normalised GPP in JULES is too low (this is a problem requiring attention in the model, probably related to canopy scaling), but for JSBACH and ORCHIDEE the GPP is approximately consistent with the observations. The observations are a little higher than the models, but this is largely influenced by underestimated LAI at Samoylov (note that for the other sites, MODIS LAI compares reasonably with ground-based estimates). Moss cover is close to 100% on Samoylov (Kutzbach et al., 2007) and by contrast, maximum LAI from MODIS is only around 0.3. This could be due to the large size of the MODIS pixels (1km$\times$1km) leading to the inclusion of water in the pixel, or because the moss has a different absorption spectrum from vascular plants and could register as bare soil. Whatever the cause, the GPP per unit LAI at Samoylov would be at least doubled by this underestimation of LAI, and if we were to account for this, the observation-based estimates would be very close to the JSBACH and ORCHIDEE results.

Aside from the low-bias in JULES, we therefore conclude that the main source of error in the modelled seasonal cycle of GPP is the huge variation in the simulated LAI. This is shown on Figure 10. For example, ORCHIDEE LAI remains at zero in the early season, when the observations and other models show carbon uptake, and it suddenly increases to a very large value later in the season, then showing an uptake that is much larger than the observations (Fig. 7). In fact, at Zackenberg the cumulative temperature is never high enough to initiate budburst in the model, so the LAI is always zero. These problems lead to unrealistic daytime emissions during spring from ORCHIDEE on Fig. 7 for most sites, and no fluxes at all for Zackenberg. Since the GPP seems to be consistent with observations when the impact of LAI is removed, we conclude that if the models could simulate the correct LAI they would largely simulate the correct GPP. JULES captures more of the difference in LAI between the sites than the other models (and subsequently captures more of the inter-site variation in GPP). This is because JULES is running a dynamic vegetation scheme that allows the vegetation fraction to vary. The LAI from JULES with fixed vegetation is also shown on Figure 10, and captures less of the inter-site variability. Therefore, both improving the LAI and including a dynamic vegetation scheme is the priority for improved simulations of tundra carbon uptake.

Carbon fluxes are also sensitive to soil moisture, as seen in simulations with increased/decreased snowfall, where differences in soil moisture availability in summer are reflected by changes in annual mean GPP, ecosystem respiration and vegetation fraction in JULES (Supplementary Figure S7), in line with Frost and Epstein (2014) . Therefore, realistic simulation of precipitation and soil moisture is a pre-requisite for improved LAI and vegetation dynamics.

3.6.2 Components of respiration

If the system were in equilibrium, the annual mean ecosystem respiration would be equal to the GPP. Thus, improving the simulation of GPP would by default improve the simulated respiration. However, the seasonal cycle of respiration is significantly different from that of GPP, due to the heterotrophic component. (This is particularly true in cold climates as the soil temperature can lag a long way behind air temperature due to the latent heat of freezing/thawing.) Furthermore, the response of respiration to changing conditions must be correctly simulated, otherwise any shift from the equilibrium state - a net source or sink of carbon - will not be correctly simulated.

It is difficult to compare the modelled respiration fluxes with the eddy covariance data (other than the annual mean). This is because the gases are assumed to be immediately emitted from the soil in the models, whereas in reality they can accumulate in the soil profile, and diffuse upwards with a significant delay. The accumulated gas may also be released from the soil in bursts, e.g. in the case of Bayelva, where the bursts of emissions in the autumn season correspond to heavy rainfall events, which (it is hypothesised) may be forcing the gas out of the soil (J. Boike, personal communication). Similarly, strong autumn emissions of CO_2 from the soil were observed by chamber measurements at Zackenberg, due to the freezing of the active layer forcing out bubbles of gas (Mastepanov et al., 2013). Further difficulty is introduced since the heterotropic and autotrophic components cannot be separated in the measurements. Therefore we cannot evaluate the soil respiration schemes in detail without direct measurements in the soil. However, one conclusion we can make is that for some models the soil carbon is approximately correct when the inputs to the system (GPP) are correct (Figure 6), which gives some indication that the decomposition models behave reasonably in these conditions.

3.6.3 Nutrient limitation and moss.

We have discussed the need for a dynamic vegetation model to capture the inter-site differences in LAI, as shown on Figure 10 where JULES using a dynamic vegetation model captures much more of the inter-site variability than the other models. However, looking more closely highlights some missing processes.

For example, the LAI at Bayelva is very small (close to zero) during the early part of the JULES simulation, but between around 2002-2006 it rapidly increases to around 1. To illustrate this transition, the fractional coverage of vegetation in JULES is shown on Supplementary Figure S3. In reality, vegetation cannot establish rapidly at a site such as this (even if climatic conditions become appropriate), because of the lack of a soil matrix and nutrients needed for plant growth, particularly nitrogen. Vascular plants could take 100's of years to establish once climatic conditions become appropriate, due to the large timescales involved in soil development. The vegetation at Bayelva is ~~mainly~~ largely mosses and lichens, which can grow in nutrient-poor conditions, but photosynthesise more slowly than vascular plants (Yuan et al., 2014). Therefore, to simulate the CO_2 flux at a very nutrient-limited site it is necessary to have a different PFT that represents the low-nutrient but low-GPP vegetation such as moss, and to include nutrient limitation for the other PFTs.

A similar problem can be seen at Samoylov, where around 90% of the site is covered by moss (Boike et al., 2013), and JULES simulates an LAI similar to that of Kytalyk (as the climatic conditions are similar), but in reality the LAI's of the two sites are very different and at Samoylov the LAI (of vascular plants) and CO_2 flux should be much smaller than that of Kytalyk. At Samoylov, the moss contributes around 40% to the total photosynthesis (Kutzbach et al., 2007), showing its importance in the carbon budget of this site. It is hypothesised that there are fewer vascular plants at Samoylov because the more waterlogged conditions (due to many polygon centre ponds) could reduce vegetation growth. In fact, reduced vegetation growth is also seen in areas with many polygon centre ponds at Kytalyk. Moreover, nitrogen may be lost in these waterlogged environments by denitrification, making it a more nutrient-limited environment.

Thus, to really capture the inter-site differences in GPP it is necessary to include nutrient limitation and other soil/plant interactions in the model. And once nutrient limitation is introduced, then moss is required (which grows in nutrient-deficient and very wet conditions where the vascular plants will not grow) in order to recreate the observed carbon uptake.

In JSBACH, moss carbon fluxes can be included - see Figure 8. This shows that the moss model can contribute significantly to the carbon budget at the mossy sites. However, at the sites with less vascular vegetation in reality (Bayelva and Samoylov), including the moss makes the total fluxes much too large, as JSBACH (like JULES) simulates too much vascular vegetation.

At Samoylov there is an early-season peak of carbon uptake that is missed in the models (Figure 7). It is possible that this could correspond to the wet ground directly following snowmelt, which leads the moss to start photosynthesising. However, it is difficult to make conclusions from the data available, and we also know that eddy covariance methods can have some problems around the time of snowmelt (for example ~~?~~ Pirk et al. (2017)). Nonetheless, we can get a clue from the moss model in JSBACH. Supplementary Figure S4 shows the annual cycle of moss GPP along with the GPP from JSBACH (without moss), showing that it captures an early-season peak before the vascular plant uptake starts in JSBACH. This plot also shows the moisture content of the moss layer, making it clear that there is a strong relationship between moisture content and moss photosynthesis. Thus it becomes even more important to simulate soil moisture correctly once moss is included in the models.

It could also be important to consider lichens separately from mosses, as their physical and biological properties can be very different. For example, the high albedo of lichens can impact the Earth's radiation budget (Bernier et al., 2011) .

4 Conclusions

Based on the analysis above, we can identify priority developments that would improve the carbon stocks and fluxes in the models. Assuming that 'state-of-the-art' is represented by a combination of the best parts of each model, we provide the following priorities for next steps to advance the state-of-the-art:

1. Improve vegetation phenology/dynamics to simulate realistic LAI (including nutrient limitation and dynamic vegetation).

2. Include moss both for photosynthesis and peat accumulation.

3. Improve the soil carbon profile for organic soils (including peat processes).

~~There~~ There is also a need to address remaining issues in the model physics, particularly for soil moisture and snow. There are feedbacks between the vegetation and the soil physical state (e.g. Sturm et al. (2001)), so incorporating more realistic vegetation such as Arctic shrubs could also lead to an improved simulation of ~~the~~ soil temperature and moisture.

There are several reasons why distinguishing between different tundra PFT's, such as grasses and shrubs, could be useful, such as differences in carbon storage, and snow interactions. Note that JULES includes a 'shrub' PFT, but these are large shrubs ($\sim$ 1.5m tall) which would not be expected to grow at the cold sites. Smaller, cold-tolerant shrubs should be added as a separate PFT. There are few modelling studies to date where tundra phenology is explicitly considered, but see Van Wijk et al. (2003) for one example.

In JSBACH the moss photosynthesis is already simulated, and the coupling to the soil carbon will be available in the next version. This provides clear guidance for other models to follow, see Porada et al. (2013, 2016). However, since JSBACH does not include nutrient limitation, the combined GPP/Reco from vascular vegetation and moss is too high (Fig. 8). Including nutrient limitation is an essential part of these priority developments.

In order to facilitate improvements to the vegetation schemes, better site-level measurements of LAI are required. This was identified as one of the largest modelling uncertainties, but only indirect satellite-derived LAI products are available, which are not sufficiently detailed or accurate for developing the model schemes. Furthermore, in order to improve the simulation of soil carbon profiles, better observations and understanding of all below-ground processes such as in-situ decomposition rates and the dynamics of cryoturbative mixing are required (Beer, 2016).

Future changes in NEE are key to understanding the role of the Arctic in a global context. We can see in Table 3 that the size of the NEE is much smaller than the errors we are currently seeing in, for example, the simulated GPP. This supports the need for the model improvements highlighted above. Future changes in the carbon balance will come both from changes in vegetation productivity/type, and decomposition of old soil carbon due to thawing permafrost. Therefore, dynamic vegetation (including nutrient limitation) is required for future simulations as well as for simulating the correct LAI in the present day. The vertical representation of soil carbon is therefore also particularly important for the fluxes in the future. However, soil carbon release will also be triggered by landscape dynamics like ground collapse and thermokarst formation, which are not yet represented in any of these models. See e.g. Schneider von Deimling et al. (2015) for a modelling study in which some of these impacts are included. This is another important aspect that must be taken into account in future model development (Rowland and Coon, 2015).

The feedbacks between the Arctic and the global climate are strongly dependent on whether carbon is released into the atmosphere from heterotrophic respiration as carbon dioxide or methane. The modelling capability at the time of this study was not sufficient to simulate the methane flux. However, this development is in progress, see e.g. Kaiser et al. (2017) , and represents an important topic for future work. Lakes and ponds also play a major role in methane and carbon dioxide exchange with the atmosphere (Bouchard et al., 2015; Langer et al., 2015) and should also be considered in future land surface models.

Accurate process representation at a site level will not necessarily transfer the same level of accuracy to a global simulation. In particular, there are issues with using a single ‘gridbox mean’ value to represent a large area of land (heterogeneity in soil/microtopography exerts non-linear controls on carbon and vegetation dynamics), and with obtaining realistic large-scale observations for quan-

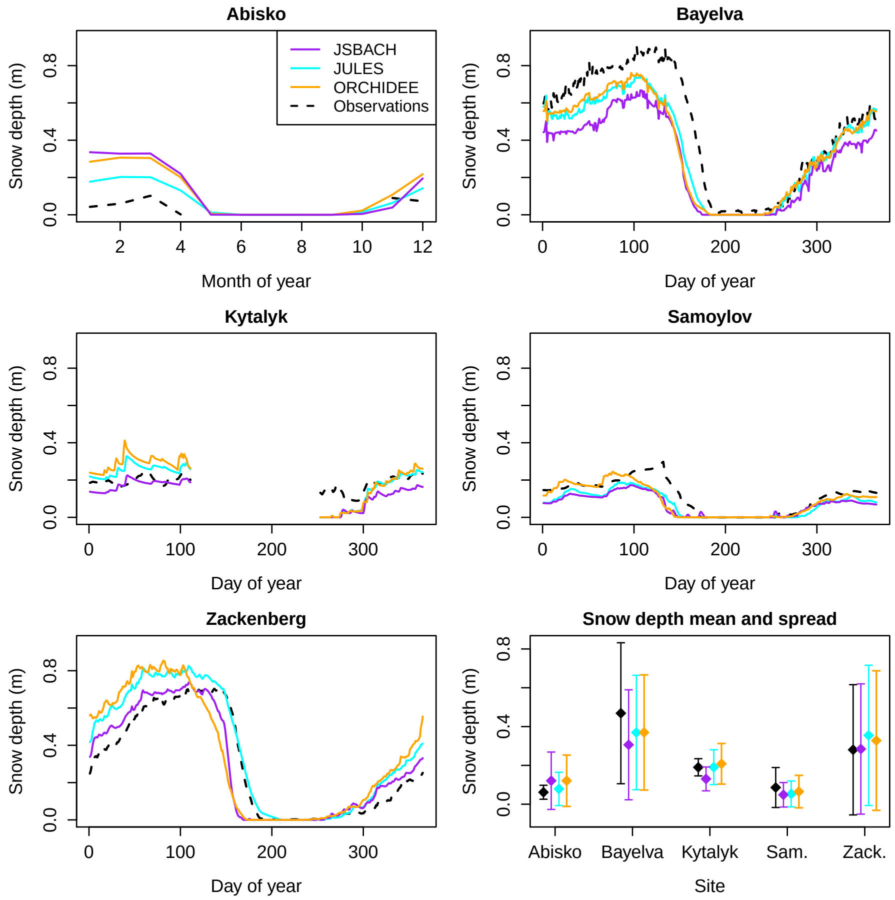


Figure 1. Mean annual cycle of snow depth at each site, showing both observations and models. On 6th panel, Samoylov and Zackenberg are abbreviated to 'Sam.' and 'Zack.'. Mean annual cycle is calculated from a single site over a number of years, except for Abisko where measurements were taken in several different locations on the mire. See supplementary Table S1 for years used at each site.

ties such as soil parameters. On the other hand, the sites used in this study represent typical tundra sites, and the model development priorities that we identify are consistent across sites, indicating that these would also lead to improved tundra carbon dynamics in global simulations. This study has allowed us to quantify deficiencies in the models that we could not have robustly identified using global datasets, due to the quantity and quality of observational data available. This work also opens up opportunities for further process studies in future.

Appendix A: Details of model set-up

Mineral soil properties were calculated from sand/silt/clay fractions. Slightly different pedotransfer functions are used in each model, but they are all taken from the same baseline soil texture (see

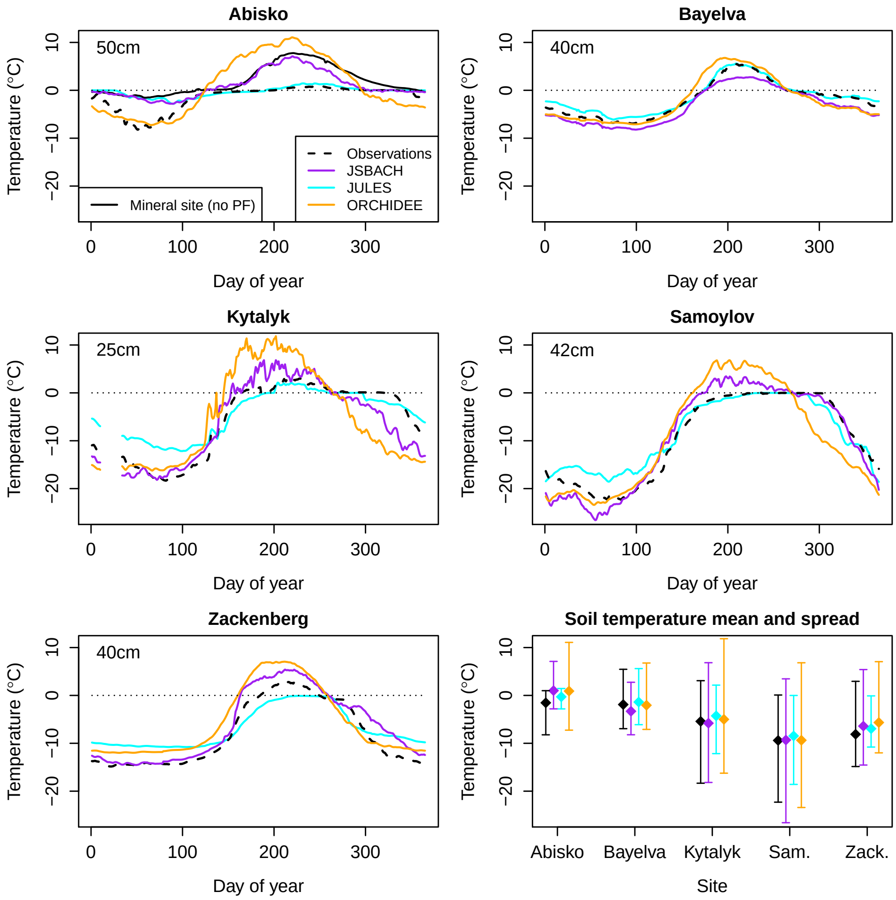


Figure 2. Mean annual cycle of soil temperature at each site, showing both observations and models. Depths of observations: Abisko: 50cm, Bayelva: 40cm, Kytalyk: 25cm, Samoylov: 42cm, Zackenberg: 40cm. JULES and ORCHIDEE take nearest soil layer and JSBACH is interpolated to correct depth, as soil layers are not well-enough resolved to get close to the right depth. On 6th panel, Samoylov and Zackenberg are abbreviated to 'Sam.' and 'Zack.'. See supplementary Table S1 for years used at each site.

Table A.1). For JULES, the organic soil fraction as a function of depth was estimated using the bulk density and carbon density. The combined organic/mineral soil properties were then calculated as in Chadburn et al. (2015a).

Assumed 'fresh' snow density for creating snowfall timeseries from snow depth: This depends on the resolution of the data. If we have low-resolution snow depth data, there may be some compaction between the snow landing and the measurement being taken, so we will use a higher density to generate the timeseries. The density used for most sites, hourly to daily resolution, is 180 kgm^{-3}. At Abisko, only 5-daily snow depth data was available, and this was at the research station rather than the mire. Since this is a relatively warm site leading to more melting, and due to the long time interval between readings, in order to give enough snow in the models a density of 240 kgm^{-3}

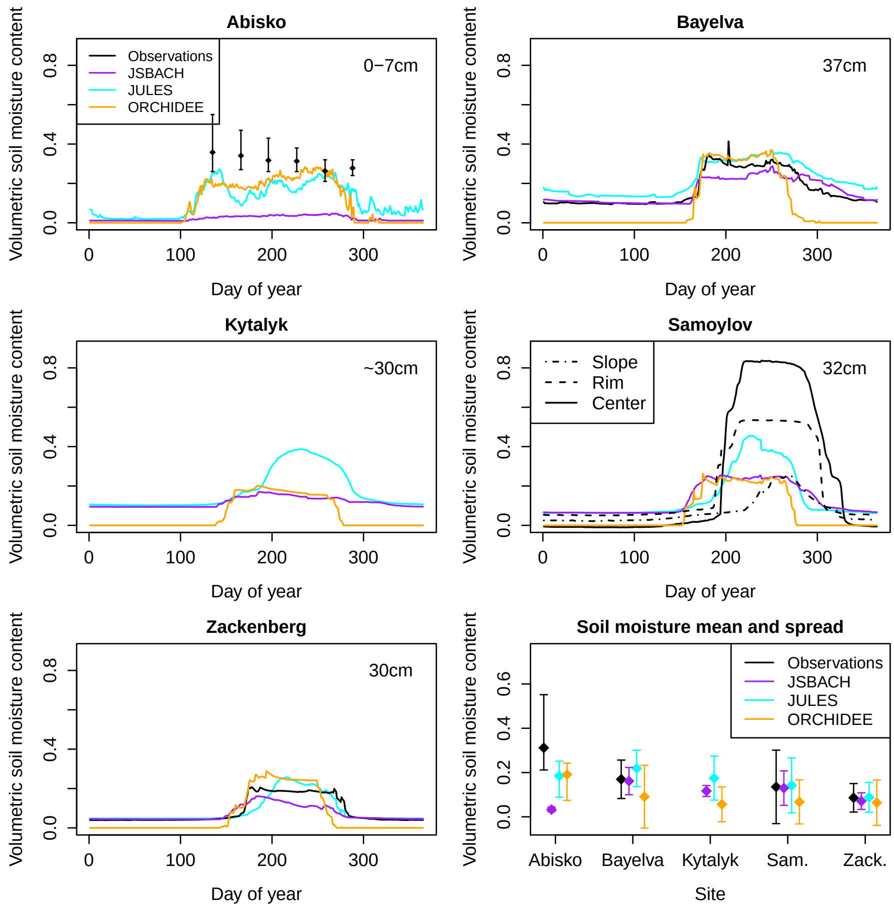


Figure 3. Mean annual cycle of unfrozen soil moisture at each site, showing both observations (where available) and models. Depths: JSBACH: 19cm for all sites (this is the closest to 30cm - the next layer is at 78cm), except Abisko, 3cm. JULES: 32cm (except Abisko, 3cm). ORCHIDEE: 36cm (except Abisko, 4cm). Observations: Bayelva: 37cm Samoylov: 32cm Zackenberg: 30cm Abisko: 0-7cm. For Samoylov, three different soil moisture profiles are shown that represent different parts of the polygonal microtopography. On 6th panel, Samoylov and Zackenberg are abbreviated to 'Sam.' and 'Zack.'. See supplementary Table S1 for years used at each site.

was used. For Abisko mire there were just a handful of snow depth measurements each year. All available values taken during a given month were averaged to give a monthly average timeseries of snow depth. We compared the depth with the model output from JULES using the forcing data prepared from the research station. The snowfall was then scaled according to the ratio of monthly snow depth in the model vs the observations. This approach introduces uncertainties that would be reduced by the availability of a higher-resolution snow depth dataset from Stordalen mire.

Acknowledgements. The authors acknowledge financial support by the European Union Seventh Framework Programme (FP7/2007-2013) project PAGE21, under GA282700. SEC, SW and GH acknowledge support

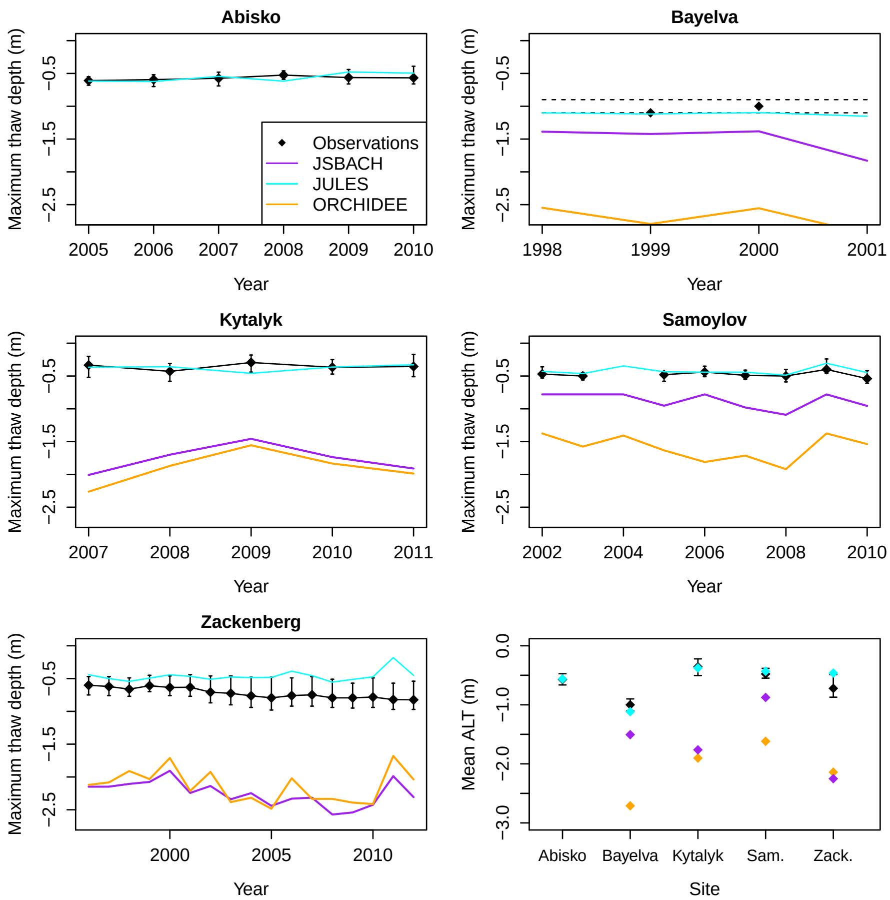


Figure 4. Maximum summer thaw depth (active layer) over a number of years at each site, comparing observations and models. Dotted lines on the second panel represent the range of observed estimates. For all other panels, CALM grids are used, and the error bars show the full range of measured values in the grid. On the final panel the error bars show the mean of the upper/lower limits from the previous panels.

from COUP (Constraining Uncertainties in Permafrost-climate Feedback) Joint Programming Initiative project (S.E.C: National Environment Research Council grant NE/M01990X/1; G.H: Swedish Research Council grant no. E0689701; S.W: Research Council of Norway project no. 244903/E10).

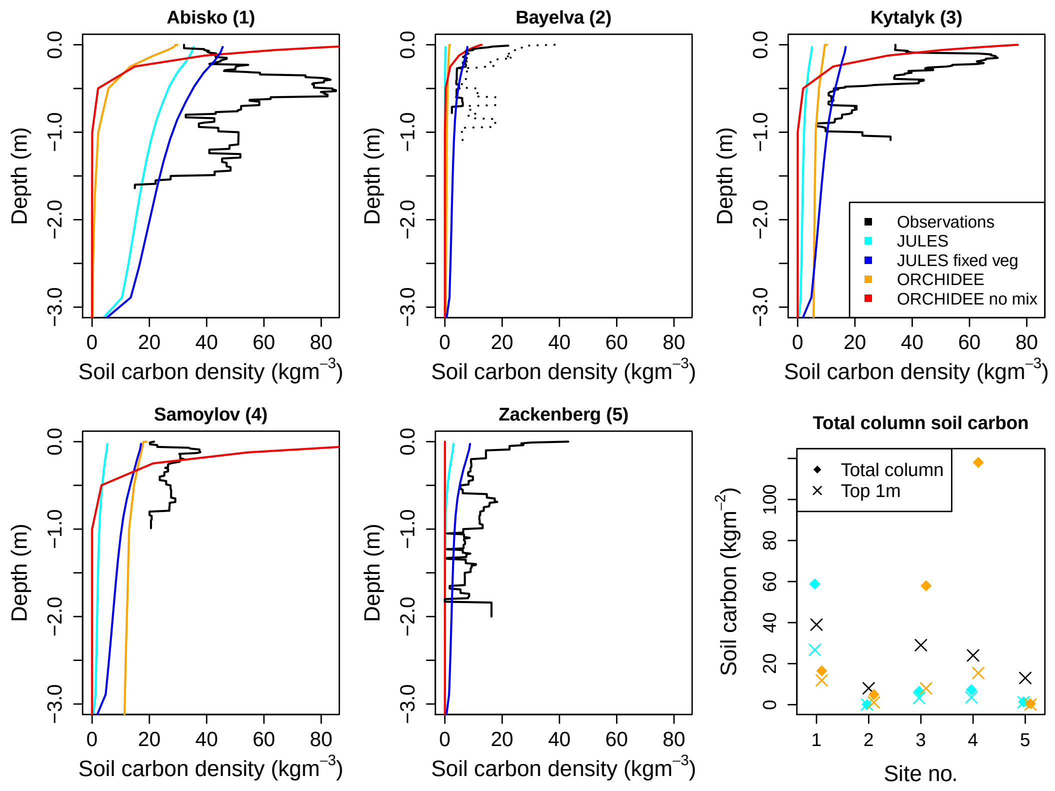


Figure 5. Profile of soil carbon at each site (kgm^{-3}). Observations and two of the models (ORCHIDEE and JULES) are shown, as these models have a vertically resolved soil carbon profile. Dotted/solid lines on second panel (Bayelva) show two different land cover types in the vicinity of the site (solid=barren ground, dotted=sparse shrub-moss tundra.) Note that site numbers on the last panel are given in the headings of the preceding panels.

References

Abermann, J., Hansen, B., Lund, M., Wacker, S., Karami, M., and Cappelen, J.: Hotspots and key periods of Greenland climate change during the past six decades, Ambio, 46, 3–11, 2017.

Åkerman, H. J. and Johansson, M.: Thawing permafrost and thicker active layers in sub-arctic Sweden, Permafrost and Periglacial Processes, 19, 279–292, 2008.

Alexandersson, H., Karlström, C., and Larsson-McCann, S.: Temperature and precipitation in Sweden 1961-90. Reference normals, Sveriges Meteorologiska och Hydrologiska Inst., 1991.

Atkin, O. K.: Reassessing the nitrogen relations of Arctic plants: a mini-review, Plant, Cell & Environment, 19, 695–704, doi:10.1111/j.1365-3040.1996.tb00404.x, http://dx.doi.org/10.1111/j.1365-3040.1996.tb00404.x, 1996.

Barr, A., Richardson, A., Hollinger, D., Papale, D., Arain, M., Black, T., Bohrer, G., Dragoni, D., Fischer, M., Gu, L., et al.: Use of change-point detection for friction–velocity threshold evaluation in eddy-covariance studies, Agricultural and Forest Meteorology, 171, 31–45, 2013.

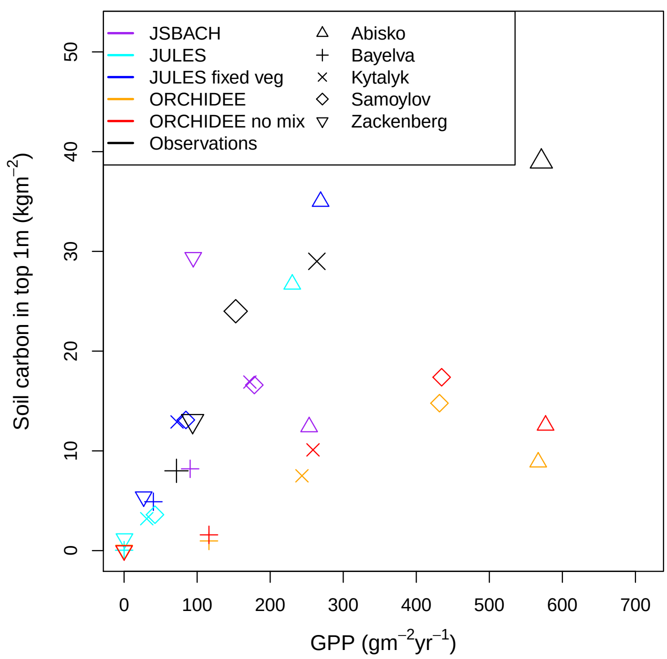


Figure 6. GPP against top 1m soil carbon at each site. The top 1m soil carbon values are for the tower footprint area (see supplementary Table S2), so that equivalent values are being compared.

Table 1. Key features of the land surface models used in this study.

Process	JSBACH	JULES	ORCHIDEE
PFT's	20	9 (+4 crop/pasture)	13
PFT that grows/is used	Tundra	C3 grass	C3 grass
Dynamic vegetation	No	Yes	No
Dynamic phenology	Yes	Yes	Yes
Nutrient limitation	No	No	No
Soil carbon	One layer	Multilayer	Multilayer
Soil carbon mixing	No	Yes	Yes
Deep soil respiration	None	Suppressed	Not suppressed
Soil latent heat	Yes	Yes	Yes
Snow	Multilayer	Multilayer	Multilayer

Bartholomeus, H., Schaepman-Strub, G., Blok, D., Sofronov, R., and Udaltsov, S.: Spectral estimation of soil properties in Siberian tundra soils and relations with plant species composition, Applied and Environmental Soil Science, 2012, 2012.

Beer, C.: Permafrost sub-grid heterogeneity of soil properties key for 3-D soil processes and future climate projections, Frontiers in Earth Science, 4, 81, 2016.

Table 2. Key climatic/physical variables at the sites.

	Abisko	Bayelva	Kytalyk	Samoylov	Zackenberg
Latitude	68.35	78.92	70.83	72.22	74.5
Longitude	19.05	11.93	147.5	126.28	-20.6
Elevation	385 m.a.s.l	25 m.a.s.l	10 m.a.s.l	6 m.a.s.l	40 m.a.s.l
Mean annual air temp.	-0.6°C	-5°C	-10.5°C	-12.5°C	-9°C
~~Summer~~ Max. monthly air temp.	11°C	5°C	10°C	10°C	6.5°C
~~Winter~~ Min. monthly air temp.	-11°C	-13°C	-34°C	-33°C	-20°C
Annual precipitation	350 mm	400 mm	230 mm	~190 mm	~~210mm~~ 260mm
Fraction as snow	~40%	~75%	~50%	~30%	~85%
Typical snow depth	0.1m	0.5-0.8m	0.2-0.4m	0.2-0.4m	0.1-1.3m
Active layer depth	0.55-1.2m	1-2m	0.25-0.5m	$<$1m	0.45-0.8m
Permafrost temperature	~0°C	-2 to -3°C	-8°C	-10°C	-6.5 to -7°C
Soil type (mineral/organic)	Organic	Mineral	Organic	Organic	Mineral

Table 3. Mean NEE budget (gCm^{-2}yr^{-1}), showing that in general this is smaller than the errors in simulated GPP, therefore the noise is larger than the signal in this data. Positive numbers represent a carbon source.

Site	JSBACH	JULES	ORCHIDEE	Observations
Abisko	-6.6	-16.0	-79.2	**-162.0**
Bayelva	-8.8	-15.1	-34.7	**-13.9**
Kytalyk	-19.0	-18.9	-24.3	**-108.0**
Samoylov	+1.5	-15.1	-58.9	**-49.6**
Zackenberg	+35.9	-5.2	+0.01	**-12.0**
Mean absolute error in GPP	*100.2*	*123.6*	*88.4*	-

Beringer, J., Lynch, A. H., Chapin, F. S., Mack, M., and Bonan, G. B.: The Representation of Arctic Soils in the Land Surface Model: The Importance of Mosses, Journal of Climate, 14, 3324–3335, doi:10.1175/1520-0442(2001)014<3324:TROASI>2.0.CO;2, http://dx.doi.org/10.1175/1520-0442(2001)014<3324:TROASI>2.0.CO;2, 2001.

Bernier, P., Desjardins, R., Karimi-Zindashty, Y., Worth, D., Beaudoin, A., Luo, Y., and Wang, S.: Boreal lichen woodlands: A possible negative feedback to climate change in eastern North America, Agricultural and Forest Meteorology, 151, 521 – 528, doi:http://dx.doi.org/10.1016/j.agrformet.2010.12.013, http://www.sciencedirect.com/science/article/pii/S0168192311000050, 2011.

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Ménard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B., and Harding, R. J.: The Joint UK Land Environment Simulator (JULES), model description Part

Table A.1. Parameters for the sites. [1]Klaminder et al. (2008). [2]J. Boike, personal communication. [3]Van Huissteden et al. (2005). [4]van der Molen et al. (2007). [5]Boike et al. (2013). [6]Hollesen et al. (2011). [7]Rydén et al. (1980). [8]Roth and Boike (2001). [9]S. Zubrzycki, soil carbon data. [10]PAGE21 Catalogue of physical parameters. [11]Bartholomeus et al. (2012). [12]Elberling et al. (2008). In most cases soil types (see Section 2.1) were translated to approx. sand/silt/clay fractions using Table 1 in Beringer et al. (2001). *Estimated from bulk density. Topographic index is from a global dataset: a 0.5° aggregate from (US Geological Survey, 2000).

	Abisko	Bayelva	Kytalyk	Samoylov	Zackenberg
~~Latitude~~	~~68.35~~	~~78.92~~	~~70.83~~	~~72.22~~	~~74.5~~
~~Longitude~~	~~19.05~~	~~11.93~~	~~147.5~~	~~126.28~~	~~-20.6~~
Organic layer thickness (cm)	∼50 [1]	0 [2]	∼20 [3,4]	∼30 [5]	5 [6]
Sand fraction	0.1	0.17	0.17	0.58	0.8 [6]
Silt fraction	0.9	0.7	0.7	0.32	0.1 [6]
Clay fraction	0.0	0.13	0.13	0.1	0.1 [6]
Bulk density	1.3 [7]	1.7 [8]	0.6 [4]	0.8 [9]	0.9-1.8 [10]
C below organic layer (kgm^{-3})	14*	0*	17 [11]*	35 [9]	10 [12]
Topographic index mean	4.0	3.9	6.2	5.9	6.7
Topographic index st.dev.	2.5	1.6	2.8	2.2	1.2

Table A.2. Soil layer thicknesses in the models.

Model	Layer thicknesses (m)
JSBACH	~~0.03, 0.19, 0.78, 2.68, 6.98, 16.44, 38.11~~ 0.06, 0.26, 0.92, 2.88, 5.72, 13.2, 30.1
JULES	0.05, 0.08, 0.11, 0.14, 0.17, 0.19, 0.22, 0.24, 0.26, 0.28, 0.30, 0.32, 0.34, 0.36, 0.38, 0.40, 0.42, 0.44, 0.46, 0.47, 0.49, 0.51, 0.53, 0.54, 0.56, 0.58, 0.59, 0.61
ORCHIDEE hydrological	0.0005, 0.002, 0.01, 0.01, 0.03, 0.06, 0.12, 0.25, 0.50, 1.00, 1.75
ORCHIDEE thermal	0.0005, 0.002, 0.01, 0.01, 0.03, 0.06, 0.12, 0.25, 0.50, 1.00, 1.75, 2.50, 3.50, 4.55, 5.66, 6.81, 8.03, 9.31, 10.65, 12.06, 13.54, 15.09, 16.72, 18.43, 20.23, 22.12, 24.10, 26.18, 28.37, 30.66, 33.07, 35.60

1: Energy and water fluxes, Geoscientific Model Development, 4, 677–699, doi:10.5194/gmd-4-677-2011, http://www.geosci-model-dev.net/4/677/2011/, 2011.

Boike, J., Roth, K., and Ippisch, O.: Seasonal snow cover on frozen ground: Energy balance calculations of a permafrost site near Ny-Ålesund, Spitsbergen, Journal of Geophysical Research: Atmospheres, 108, 2003.

Boike, J., Ippisch, O., Overduin, P. P., Hagedorn, B., and Roth, K.: Water, heat and solute dynamics of a mud boil, Spitsbergen, Geomorphology, 95, 61–73, 2008a.

Boike, J., Wille, C., and Abnizova, A.: Climatology and summer energy and water balance of polygonal tundra in the Lena River Delta, Siberia, Journal of Geophysical Research: Biogeosciences, 113, 2008b.

Boike, J., Kattenstroth, B., Abramova, K., Bornemann, N., Chetverova, A., Fedorova, I., Fröb, K., Grigoriev, M., Grüber, M., Kutzbach, L., Langer, M., Minke, M., Muster, S., Piel, K., Pfeiffer, E.-M., Stoof, G., Westermann, S., Wischnewski, K., Wille, C., and Hubberten, H.-W.: Baseline characteristics of climate, permafrost and land cover from a new permafrost observatory in the Lena River Delta, Siberia (1998-2011), Biogeosciences, 10, 2105–2128, doi:10.5194/bg-10-2105-2013, http://www.biogeosciences.net/10/2105/2013/, 2013.

Bouchard, F., Laurion, I., Prėskienis, V., Fortier, D., Xu, X., and Whiticar, M. J.: Modern to millennium-old greenhouse gases emitted from ponds and lakes of the Eastern Canadian Arctic (Bylot Island, Nunavut), Biogeosciences, 12, 7279–7298, doi:10.5194/bg-12-7279-2015, https://www.biogeosciences.net/12/7279/2015/, 2015.

Brovkin, V., Raddatz, T., Reick, C. H., Claussen, M., and Gayler, V.: Global biogeophysical interactions between forest and climate, Geophysical Research Letters, 36, 2009.

Brown, J., Ferrians Jr, O. J., Heginbottom, J., and Melnikov, E.: Circum-arctic map of permafrost and ground ice conditions, National Snow and Ice Data Center, http://nsidc.org/data/docs/fgdc/ggd318_map_circumarctic, 1998.

Burke, E. J., Hartley, I. P., and Jones, C. D.: Uncertainties in the global temperature change caused by carbon release from permafrost thawing, The Cryosphere, 6, 1063–1076, doi:10.5194/tc-6-1063-2012, http://www.the-cryosphere.net/6/1063/2012/, 2012.

Burke, E. J., Jones, C. D., and Koven, C. D.: Estimating the permafrost-carbon climate response in the CMIP5 climate models using a simplified approach, Journal of Climate, 26, 4897–4909, 2013.

Burke, E. J., Chadburn, S. E., and Ekici, A.: A vertical representation of soil carbon in the JULES land surface scheme (vn4.3_permafrost) with a focus on permafrost regions, Geoscientific Model Development, 10, 959–975, doi:10.5194/gmd-10-959-2017, http://www.geosci-model-dev.net/10/959/2017/, 2017.

Cahoon, S. M. P., Sullivan, P. F., Shaver, G. R., Welker, J. M., and Post, E.: Interactions among shrub cover and the soil microclimate may determine future Arctic carbon budgets, Ecology Letters, 15, 1415–1422, doi:10.1111/j.1461-0248.2012.01865.x, http://dx.doi.org/10.1111/j.1461-0248.2012.01865.x, 2012.

Callaghan, T. V., Bergholm, F., Christensen, T. R., Jonasson, C., Kokfelt, U., and Johansson, M.: A new climate era in the sub-Arctic: Accelerating climate changes and multiple impacts, Geophysical Research Letters, 37, 2010.

Cannone, N., Augusti, A., Malfasi, F., Pallozzi, E., Calfapietra, C., and Brugnoli, E.: The interaction of biotic and abiotic factors at multiple spatial scales affects the variability of CO2 fluxes in polar environments, Polar Biology, 39, 1581–1596, 2016.

Carlsson, B. A., Karlsson, P. S., and Svensson, B. M.: Alpine and subalpine vegetation, Acta Phytogeographica Suecica, 84, 75–90, 1999.

Chadburn, S., Burke, E., Essery, R., Boike, J., Langer, M., Heikenfeld, M., Cox, P., and Friedlingstein, P.: An improved representation of physical permafrost dynamics in the JULES land-surface model, Geoscientific Model Development, 8, 1493–1508, doi:10.5194/gmd-8-1493-2015, http://www.geosci-model-dev.net/8/1493/2015/, 2015a.

Chadburn, S. E., Burke, E. J., Essery, R. L. H., Boike, J., Langer, M., Heikenfeld, M., Cox, P. M., and Friedlingstein, P.: Impact of model developments on present and future simulations of permafrost in a global land-surface model, The Cryosphere, 9, 1505–1521, doi:10.5194/tc-9-1505-2015, http://www.the-cryosphere.net/9/1505/2015/, 2015b.

Christiansen, H. H., Etzelmüller, B., Isaksen, K., Juliussen, H., Farbrot, H., Humlum, O., Johansson, M., Ingeman-Nielsen, T., Kristensen, L., Hjort, J., et al.: The thermal state of permafrost in the Nordic area during the International Polar Year 2007–2009, Permafrost and Periglacial Processes, 21, 156–181, 2010.

Clark, D. B., Mercado, L. M., Sitch, S., Jones, C. D., Gedney, N., Best, M. J., Pryor, M., Rooney, G. G., Essery, R. L. H., Blyth, E., Boucher, O., Harding, R. J., Huntingford, C., and Cox, P. M.: The Joint UK Land Environment Simulator (JULES), model description Part 2: Carbon fluxes and vegetation dynamics, Geoscientific Model Development, 4, 701–722, doi:10.5194/gmd-4-701-2011, http://www.geosci-model-dev.net/4/701/2011/, 2011.

Cohen, W. B., Maiersperger, T. K., Turner, D. P., Ritts, W. D., Pflugmacher, D., Kennedy, R. E., Kirschbaum, A., Running, S. W., Costa, M., and Gower, S. T.: MODIS land cover and LAI collection 4 Product quality across nine sites in the western hemisphere, IEEE Transactions on Geoscience and Remote Sensing, 44, 1843–1857, 2006.

Collatz, G. J., Ribas-Carbo, M., and Berry, J.: Coupled photosynthesis-stomatal conductance model for leaves of C4 plants, Functional Plant Biology, 19, 519–538, 1992.

Cristóbal, J., Prakash, A., Anderson, M. C., Kustas, W. P., Euskirchen, E. S., and Kane, D. L.: Estimation of surface energy fluxes in the Arctic tundra using the remote sensing thermal-based Two-Source Energy Balance model, Hydrology and Earth System Sciences, 21, 1339–1358, doi:10.5194/hess-21-1339-2017, http://www.hydrol-earth-syst-sci.net/21/1339/2017/, 2017.

Domine, F., Barrere, M., and Sarrazin, D.: Seasonal evolution of the effective thermal conductivity of the snow and the soil in high Arctic herb tundra at Bylot Island, Canada, The Cryosphere, 10, 2573–2588, doi:10.5194/tc-10-2573-2016, https://www.the-cryosphere.net/10/2573/2016/, 2016.

Dyrness, C.: Control of depth to permafrost and soil temperature by the forest floor in black spruce/feathermoss communities, Pacific Northwest Forest and Range Experiment Station (Portland, Or.), 396, 1982.

Ekici, A., Beer, C., Hagemann, S., Boike, J., Langer, M., and Hauck, C.: Simulating high-latitude permafrost regions by the JSBACH terrestrial ecosystem model, Geoscientific Model Development, 7, 631–647, doi:10.5194/gmd-7-631-2014, http://www.geosci-model-dev.net/7/631/2014/, 2014.

Ekici, A., Chadburn, S., Chaudhary, N., Hajdu, L., Marmy, A., Peng, S., Boike, J., Burke, E., Friend, A., Hauck, C., et al.: Site-level model intercomparison of high latitude and high altitude soil thermal dynamics in tundra and barren landscapes, The Cryosphere, 9, 1343–1361, 2015.

Elberling, B., Tamstorf, M. P., Michelsen, A., Arndal, M. F., Sigsgaard, C., Illeris, L., Bay, C., Hansen, B. U., Christensen, T. R., Hansen, E. S., et al.: Soil and plant community-characteristics and dynamics at Zackenberg, Advances in Ecological Research, 40, 223–248, 2008.

Elberling, B., Michelsen, A., Schädel, C., Schuur, E. A., Christiansen, H. H., Berg, L., Tamstorf, M. P., and Sigsgaard, C.: Long-term CO2 production following permafrost thaw, Nature Climate Change, 3, 890–894, 2013.

Epstein, H., Bhatt, U., Raynolds, M., Walker, D., Forbes, B., Macias-Fauria, M., Loranty, M., Phoenix, G., and Bjerke, J.: Tundra Greenness, www. arctic. noaa. gov/reportcard, p. 59, 2016.

Farquhar, G. v., Caemmerer, S. v., and Berry, J.: A biochemical model of photosynthetic CO 2 assimilation in leaves of C 3 species, Planta, 149, 78–90, 1980.

Fedorova, I., Chetverova, A., Bolshiyanov, D., Makarov, A., Boike, J., Heim, B., Morgenstern, A., Overduin, P. P., Wegner, C., Kashina, V., et al.: Lena Delta hydrology and geochemistry: long-term hydrological data and recent field observations, Biogeosciences, 12, 345–363, 2015.

Førland, E. J., Benestad, R., Hanssen-Bauer, I., Haugen, J. E., and Skaugen, T. E.: Temperature and precipitation development at Svalbard 1900–2100, Advances in Meteorology, 2011, 2012.

Frolking, S., Roulet, N. T., Tuittila, E., Bubier, J. L., Quillet, A., Talbot, J., and Richard, P. J. H.: A new model of Holocene peatland net primary production, decomposition, water balance, and peat accumulation, Earth System Dynamics, 1, 1–21, doi:10.5194/esd-1-1-2010, http://www.earth-syst-dynam.net/1/1/2010/, 2010.

Frost, G. V. and Epstein, H. E.: Tall shrub and tree expansion in Siberian tundra ecotones since the 1960s, Global Change Biology, 20, 1264–1277, doi:10.1111/gcb.12406, http://dx.doi.org/10.1111/gcb.12406, 2014.

Gavrilov, A., Romanovskii, N., Romanovsky, V., Hubberten, H.-W., and Tumskoy, V.: Reconstruction of ice complex remnants on the eastern Siberian Arctic Shelf, Permafrost and Periglacial Processes, 14, 187–198, 2003.

Gedney, N. and Cox, P. M.: The Sensitivity of Global Climate Model Simulations to the Representation of Soil Moisture, Journal of Hydrometeorology, 4, 1265–1275, doi:10.1175/1525-7541(2003)004<1265:TSOGCM>2.0.CO;2, http://dx.doi.org/10.1175/1525-7541(2003)004<1265:TSOGCM>2.0.CO;2, 2003.

Goll, D. S., Brovkin, V., Liski, J., Raddatz, T., Thum, T., and Todd-Brown, K. E.: Strong dependence of CO2 emissions from anthropogenic land cover change on initial land cover and soil carbon parametrization, Global Biogeochemical Cycles, 29, 1511–1523, 2015.

Gouttevin, I., Krinner, G., Ciais, P., Polcher, J., and Legout, C.: Multi-scale validation of a new soil freezing scheme for a land-surface model with physically-based hydrology, The Cryosphere, 6, 407–430, doi:10.5194/tc-6-407-2012, http://www.the-cryosphere.net/6/407/2012/, 2012a.

Gouttevin, I., Menegoz, M., Dominé, F., Krinner, G., Koven, C., Ciais, P., Tarnocai, C., and Boike, J.: How the insulating properties of snow affect soil carbon distribution in the continental pan-Arctic area, Journal of Geophysical Research: Biogeosciences, 117, n/a–n/a, doi:10.1029/2011JG001916, http://dx.doi.org/10.1029/2011JG001916, 2012b.

Hagemann, S. and Stacke, T.: Impact of the soil hydrology scheme on simulated soil moisture memory, Climate Dynamics, 44, 1731–1750, 2015.

Hanssen-Bauer, I. and Førland, E.: Long-term trends in precipitation and temperature in the Norwegian Arctic: can they be explained by changes in atmospheric circulation patterns?, Climate Research, 10, 143–153, 1998.

Harper, A. B., Cox, P. M., Friedlingstein, P., Wiltshire, A. J., Jones, C. D., Sitch, S., Mercado, L. M., Groenendijk, M., Robertson, E., Kattge, J., Bönisch, G., Atkin, O. K., Bahn, M., Cornelissen, J., Niinemets, U., Onipchenko, V., Peñuelas, J., Poorter, L., Reich, P. B., Soudzilovskaia, N. A., and Bodegom, P. V.: Improved representation of plant functional types and physiology in the Joint UK Land Environment Simulator (JULES v4.2) using plant trait information, Geoscientific Model Development, 9, 2415–2440, doi:10.5194/gmd-9-2415-2016, http://www.geosci-model-dev.net/9/2415/2016/, 2016.

Hayes, D. J., McGuire, A. D., Kicklighter, D. W., Gurney, K. R., Burnside, T. J., and Melillo, J. M.: Is the northern high-latitude land-based CO2 sink weakening?, Global Biogeochemical Cycles, 25, n/a–n/a, doi:10.1029/2010GB003813, http://dx.doi.org/10.1029/2010GB003813, gB3018, 2011.

Hollesen, J., Elberling, B., and Jansson, P.-E.: Future active layer dynamics and carbon dioxide production from thawing permafrost layers in Northeast Greenland, Global Change Biology, 17, 911–926, 2011.

Hugelius, G., Strauss, J., Zubrzycki, S., Harden, J. W., Schuur, E. A. G., Ping, C.-L., Schirrmeister, L., Grosse, G., Michaelson, G. J., Koven, C. D., O'Donnell, J. A., Elberling, B., Mishra, U., Camill, P., Yu, Z., Palmtag, J., and Kuhry, P.: Estimated stocks of circumpolar permafrost carbon with quantified uncertainty ranges and identified data gaps, Biogeosciences, 11, 6573–6593, doi:10.5194/bg-11-6573-2014, http://www.biogeosciences.net/11/6573/2014/, 2014.

Jammet, M., Crill, P., Dengel, S., and Friborg, T.: Large methane emissions from a subarctic lake during spring thaw: Mechanisms and landscape significance, Journal of Geophysical Research: Biogeosciences, 120, 2289–2305, 2015.

Jammet, M., Dengel, S., Kettner, E., Parmentier, F.-J. W., Wik, M., Crill, P., and Friborg, T.: Year-round CH_4 and CO_2 flux dynamics in two contrasting freshwater ecosystems of the subarctic, Biogeosciences Discussions, 2017, 1–49, doi:10.5194/bg-2016-466, http://www.biogeosciences-discuss.net/bg-2016-466/, 2017.

Johansson, M., Christensen, T. R., Akerman, H. J., and Callaghan, T. V.: What determines the current presence or absence of permafrost in the Torneträsk Region, a sub-Arctic landscape in Northern Sweden?, AMBIO: A Journal of the Human Environment, 35, 190–197, 2006.

Johansson, M., Åkerman, J., Keuper, F., Christensen, T. R., Lantuit, H., and Callaghan, T. V.: Past and present permafrost temperatures in the Abisko area: Redrilling of boreholes, Ambio, 40, 558, 2011.

Johansson, M., Callaghan, T. V., Bosiö, J., Åkerman, H. J., Jackowicz-Korczynski, M., and Christensen, T. R.: Rapid responses of permafrost and vegetation to experimentally increased snow cover in sub-arctic Sweden, Environmental Research Letters, 8, 035 025, 2013.

Jones, C. and Sellar, A.: Development of the 1st version of the UK Earth system model, UKESM newsletter no.1 - August 2015, http://www.jwcrp.org.uk/research-activity/ukesm-newsarchive.asp, 2015.

Kaiser, S., Göckede, M., Castro-Morales, K., Knoblauch, C., Ekici, A., Kleinen, T., Zubrzycki, S., Sachs, T., Wille, C., and Beer, C.: Process-based modelling of the methane balance in periglacial landscapes (JSBACH-methane), Geoscientific Model Development, 10, 333–358, doi:10.5194/gmd-10-333-2017, https://www.geosci-model-dev.net/10/333/2017/, 2017.

Klaminder, J., Yoo, K., Rydberg, J., and Giesler, R.: An explorative study of mercury export from a thawing palsa mire, Journal of Geophysical Research: Biogeosciences, 113, 2008.

Knorr, W.: Annual and interannual CO2 exchanges of the terrestrial biosphere: Process-based simulations and uncertainties, Global Ecology and Biogeography, 9, 225–252, 2000.

Koven, C., Friedlingstein, P., Ciais, P., Khvorostyanov, D., Krinner, G., and Tarnocai, C.: On the formation of high-latitude soil carbon stocks: Effects of cryoturbation and insulation by organic matter in a land surface model, Geophysical Research Letters, 36, n/a–n/a, doi:10.1029/2009GL040150, http://dx.doi.org/10.1029/2009GL040150, 2009.

Koven, C. D., Riley, W. J., Subin, Z. M., Tang, J. Y., Torn, M. S., Collins, W. D., Bonan, G. B., Lawrence, D. M., and Swenson, S. C.: The effect of vertically resolved soil biogeochemistry and alternate soil C and

N models on C dynamics of CLM4, Biogeosciences, 10, 7109–7131, doi:10.5194/bg-10-7109-2013, http://www.biogeosciences.net/10/7109/2013/, 2013.

Koven, C. D., Schuur, E., Schädel, C., Bohn, T., Burke, E., Chen, G., Chen, X., Ciais, P., Grosse, G., Harden, J. W., et al.: A simplified, data-constrained approach to estimate the permafrost carbon–climate feedback, Phil. Trans. R. Soc. A, 373, 20140 423, 2015.

Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., and Prentice, I. C.: A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system, Global Biogeochemical Cycles, 19, n/a–n/a, doi:10.1029/2003GB002199, http://dx.doi.org/10.1029/2003GB002199, gB1015, 2005.

Kutzbach, L., Wille, C., and Pfeiffer, E.-M.: The exchange of carbon dioxide between wet arctic tundra and the atmosphere at the Lena River Delta, Northern Siberia, Biogeosciences, 4, 869–890, 2007.

Langer, M., Westermann, S., Muster, S., Piel, K., and Boike, J.: The surface energy balance of a polygonal tundra site in northern Siberia - Part 1: Spring to fall, The Cryosphere, 5, 151–171, doi:10.5194/tc-5-151-2011, http://www.the-cryosphere.net/5/151/2011/, 2011a.

Langer, M., Westermann, S., Muster, S., Piel, K., and Boike, J.: The surface energy balance of a polygonal tundra site in northern Siberia - Part 2: Winter, The Cryosphere, 5, 509–524, doi:10.5194/tc-5-509-2011, http://www.the-cryosphere.net/5/509/2011/, 2011b.

Langer, M., Westermann, S., Anthony, K. W., Wischnewski, K., and Boike, J.: Frozen ponds: production and storage of methane during the Arctic winter in a lowland tundra landscape in northern Siberia, Lena River delta, Biogeosciences, 12, 977, 2015.

Lantmäteriet: Vegetation map of Abisko 30I and Rensjön 30J, Produced and digitised by Lantmäteriet GSD, Scale 1:100 000, 1997.

Lasslop, G., Reichstein, M., Papale, D., Richardson, A. D., Arneth, A., Barr, A., Stoy, P., and Wohlfahrt, G.: Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation, Global Change Biology, 16, 187–208, 2010.

Lawrence, D. and Slater, A.: Incorporating organic soil into a global climate model, Climate Dynamics, 30, 145–160, doi:10.1007/s00382-007-0278-1, http://dx.doi.org/10.1007/s00382-007-0278-1, 2008.

López-Moreno, J., Boike, J., Sanchez-Lorenzo, A., and Pomeroy, J.: Impact of climate warming on snow processes in Ny-Ålesund, a polar maritime site at Svalbard, Global and Planetary Change, 146, 10–21, 2016.

Lüers, J., Westermann, S., Piel, K., and Boike, J.: Annual CO_2 budget and seasonal CO_2 exchange signals at a high Arctic permafrost site on Spitsbergen, Svalbard archipelago, Biogeosciences, 11, 6307–6322, 2014.

Lund, M., Falk, J. M., Friborg, T., Mbufong, H. N., Sigsgaard, C., Soegaard, H., and Tamstorf, M. P.: Trends in CO_2 exchange in a high Arctic tundra heath, 2000–2010, Journal of Geophysical Research: Biogeosciences, 117, 2012.

Lund, M., Hansen, B. U., Pedersen, S. H., Stiegler, C., and Tamstorf, M. P.: Characteristics of summer-time energy exchange in a high Arctic tundra heath 2000-2010, Tellus B, 66, 2014.

Lund, M., Stiegler, C., Abermann, J., Citterio, M., Hansen, B. U., and van As, D.: Spatiotemporal variability in surface energy balance across tundra, snow and ice in Greenland, Ambio, 46, 81–93, 2017.

MacDougall, A. H. and Knutti, R.: Projecting the release of carbon from permafrost soils using a perturbed parameter ensemble modelling approach, Biogeosciences, 13, 2123–2136, doi:10.5194/bg-13-2123-2016, http://www.biogeosciences.net/13/2123/2016/, 2016.

Marcott, S. A., Shakun, J. D., Clark, P. U., and Mix, A. C.: A reconstruction of regional and global temperature for the past 11,300 years, science, 339, 1198–1201, 2013.

Mastepanov, M., Sigsgaard, C., Dlugokencky, E. J., Houweling, S., Ström, L., Tamstorf, M. P., and Christensen, T. R.: Large tundra methane burst during onset of freezing, Nature, 456, 628–630, 2008.

Mastepanov, M., Sigsgaard, C., Tagesson, T., Ström, L., Tamstorf, M. P., Lund, M., and Christensen, T. R.: Revisiting factors controlling methane emissions from high-Arctic tundra, Biogeosciences, 10, 5139–5158, doi:10.5194/bg-10-5139-2013, http://www.biogeosciences.net/10/5139/2013/, 2013.

Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M. L. T., Lamarque, J.-F., Matsumoto, K., Montzka, S. A., Raper, S. C. B., Riahi, K., Thomson, A., Velders, G. J. M., and van Vuuren, D. P.: The RCP greenhouse gas concentrations and their extensions from 1765 to 2300, Climatic Change, 109, 213–241, doi:10.1007/s10584-011-0156-z, http://dx.doi.org/10.1007/s10584-011-0156-z, 2011.

MODIS15A2: NASA EOSDIS Land Processes DAAC, USGS Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota (https://lpdaac.usgs.gov), accessed [12 13, 2016].

Morgenstern, A., Ulrich, M., Günther, F., Roessler, S., Fedorova, I. V., Rudaya, N. A., Wetterich, S., Boike, J., and Schirrmeister, L.: Evolution of thermokarst in East Siberian ice-rich permafrost: A case study, Geomorphology, 201, 363–379, 2013.

Muster, S., Langer, M., Heim, B., Westermann, S., and Boike, J.: Subpixel heterogeneity of ice-wedge polygonal tundra: a multi-scale analysis of land cover and evapotranspiration in the Lena River Delta, Siberia, Tellus B, 64, http://www.tellusb.net/index.php/tellusb/article/view/17301, 2012.

Ohtsuka, T., Adachi, M., Uchida, M., and Nakatsubo, T.: Relationships between vegetation types and soil properties along a topographical gradient on the northern coast of the Brgger Peninsula, Svalbard, Polar bioscience, 19, 63–72, 2006.

Palmtag, J., Hugelius, G., Lashchinskiy, N., Tamstorf, M. P., Richter, A., Elberling, B., and Kuhry, P.: Storage, Landscape Distribution, and Burial History of Soil Organic Matter in Contrasting Areas of Continuous Permafrost, Arctic, Antarctic, and Alpine Research, 47, 71–88, doi:10.1657/AAAR0014-027, http://aaarjournal.org/doi/abs/10.1657/AAAR0014-027, 2015.

Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Kutsch, W., Longdoz, B., Rambal, S., Valentini, R., Vesala, T., et al.: Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation, Biogeosciences, 3, 571–583, 2006.

Parmentier, F., Van Der Molen, M., Van Huissteden, J., Karsanaev, S., Kononov, A., Suzdalov, D., Maximov, T., and Dolman, A.: Longer growing seasons do not increase net carbon uptake in the northeastern Siberian tundra, Journal of Geophysical Research: Biogeosciences, 116, 2011.

Pedersen, E., Elberling, B., and Michelsen, A.: Seasonal variations in methane fluxes in response to summer warming and leaf litter addition in a subarctic heath ecosystem., Journal of Geophysical Research: Biogeosciences (revised)., 2017.

Pedersen, S. H., Tamstorf, M. P., Abermann, J., Westergaard-Nielsen, A., Lund, M., Skov, K., Sigsgaard, C., Mylius, M. R., Hansen, B. U., Liston, G. E., et al.: Spatiotemporal characteristics of seasonal snow cover in Northeast Greenland from in situ observations, Arctic, Antarctic, and Alpine Research, 48, 653–671, 2016.

Peters, G. P., Andrew, R. M., Solomon, S., and Friedlingstein, P.: Measuring a fair and ambitious climate agreement using cumulative emissions, Environmental Research Letters, 10, 105 004, http://stacks.iop.org/1748-9326/10/i=10/a=105004, 2015.

Pirk, N., Sievers, J., Mertes, J., Parmentier, F.-J. W., Mastepanov, M., and Christensen, T. R.: Spatial variability of CO_2 uptake in polygonal tundra: assessing low-frequency disturbances in eddy covariance flux estimates, Biogeosciences, 14, 3157–3169, doi:10.5194/bg-14-3157-2017, https://www.biogeosciences.net/14/3157/2017/, 2017.

Porada, P., Weber, B., Elbert, W., Pöschl, U., and Kleidon, A.: Estimating global carbon uptake by lichens and bryophytes with a process-based model, Biogeosciences, 10, 6989–7033, doi:10.5194/bg-10-6989-2013, http://www.biogeosciences.net/10/6989/2013/, 2013.

Porada, P., Ekici, A., and Beer, C.: Effects of bryophyte and lichen cover on permafrost soil temperature at large scale, The Cryosphere, 10, 2291–2315, doi:10.5194/tc-10-2291-2016, https://www.the-cryosphere.net/10/2291/2016/, 2016.

Qian, H., Joseph, R., and Zeng, N.: Enhanced terrestrial carbon uptake in the Northern High Latitudes in the 21st century from the Coupled Carbon Cycle Climate Model Intercomparison Project model projections, Global Change Biology, 16, 641–656, doi:10.1111/j.1365-2486.2009.01989.x, http://dx.doi.org/10.1111/j.1365-2486.2009.01989.x, 2010.

Quegan, S., Beer, C., Shvidenko, A., McCallum, I., Handoh, I. C., Peylin, P., Roedenbeck, C., Lucht, W., Nilsson, S., and Schmullius, C.: Estimating the carbon balance of central Siberia using a landscape-ecosystem approach, atmospheric inversion and Dynamic Global Vegetation Models, Global Change Biology, 17, 351–365, 2011.

Raddatz, T., Reick, C., Knorr, W., Kattge, J., Roeckner, E., Schnur, R., Schnitzler, K.-G., Wetzel, P., and Jungclaus, J.: Will the tropical land biosphere dominate the climate–carbon cycle feedback during the twenty-first century?, Climate Dynamics, 29, 565–574, 2007.

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., et al.: On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm, Global Change Biology, 11, 1424–1439, 2005.

Ridefelt, H., Etzelmüller, B., Boelhouwers, J., and Jonasson, C.: Statistic-empirical modelling of mountain permafrost distribution in the Abisko region, sub-Arctic northern Sweden, Norsk Geografisk Tidsskrift-Norwegian Journal of Geography, 62, 278–289, 2008.

Roth, K. and Boike, J.: Quantifying the thermal dynamics of a permafrost site near Ny-Ålesund, Svalbard, Water resources research, 37, 2901–2914, 2001.

Rowland, J. C. and Coon, E. T.: From documentation to prediction: raising the bar for thermokarst research, Hydrogeology Journal, 24, 2015.

Rydén, B., Fors, L., and Kostov, L.: Physical properties of the tundra soil-water system at Stordalen, Abisko, Ecological Bulletins, pp. 27–54, 1980.

Schaphoff, S., Heyder, U., Ostberg, S., Gerten, D., Heinke, J., and Lucht, W.: Contribution of permafrost soils to the global carbon budget, Environmental Research Letters, 8, 014 026, http://stacks.iop.org/1748-9326/8/i=1/a=014026, 2013.

Schirrmeister, L., Siegert, C., Kuznetsova, T., Kuzmina, S., Andreev, A., Kienast, F., Meyer, H., and Bobrov, A.: Paleoenvironmental and paleoclimatic records from permafrost deposits in the Arctic region of Northern Siberia, Quaternary International, 89, 97–118, 2002.

Schneider von Deimling, T., Meinshausen, M., Levermann, A., Huber, V., Frieler, K., Lawrence, D. M., and Brovkin, V.: Estimating the near-surface permafrost-carbon feedback on global warming, Biogeosciences, 9, 649–665, doi:10.5194/bg-9-649-2012, http://www.biogeosciences.net/9/649/2012/, 2012.

Schneider von Deimling, T., Grosse, G., Strauss, J., Schirrmeister, L., Morgenstern, A., Schaphoff, S., Meinshausen, M., and Boike, J.: Observation-based modelling of permafrost carbon fluxes with accounting for deep carbon deposits and thermokarst activity, Biogeosciences, 12, 3469–3488, doi:10.5194/bg-12-3469-2015, http://www.biogeosciences.net/12/3469/2015/, 2015.

Schuldt, R. J., Brovkin, V., Kleinen, T., and Winderlich, J.: Modelling Holocene carbon accumulation and methane emissions of boreal wetlands - an Earth system model approach, Biogeosciences, 10, 1659–1674, doi:10.5194/bg-10-1659-2013, http://www.biogeosciences.net/10/1659/2013/, 2013.

Schuur, E. A. G., McGuire, A. D., Schadel, C., Grosse, G., Harden, J. W., Hayes, D. J., Hugelius, G., Koven, C. D., Kuhry, P., Lawrence, D. M., Natali, S. M., Olefeldt, D., Romanovsky, V. E., Schaefer, K., Turetsky, M. R., Treat, C. C., and Vonk, J. E.: Climate change and the permafrost carbon feedback, Nature, 520, 171–179, doi:10.1038/nature14338, http://dx.doi.org/10.1038/nature14338, 2015.

Schwamborn, G., Rachold, V., and Grigoriev, M. N.: Late Quaternary sedimentation history of the Lena Delta, Quaternary international, 89, 119–134, 2002.

Siewert, M. B., Hanisch, J., Weiss, N., Kuhry, P., Maximov, T. C., and Hugelius, G.: Comparing carbon storage of Siberian tundra and taiga permafrost ecosystems at very high spatial resolution, Journal of Geophysical Research: Biogeosciences, 120, 1973–1994, doi:10.1002/2015JG002999, http://dx.doi.org/10.1002/2015JG002999, 2015JG002999, 2015.

Siewert, M. B., Hugelius, G., Heim, B., and Faucherre, S.: Landscape controls and vertical variability of soil organic carbon storage in permafrost-affected soils of the Lena River Delta, Catena, 147, 725–741, 2016.

Stiegler, C., Lund, M., Christensen, T. R., Mastepanov, M., and Lindroth, A.: Two years with extreme and little snowfall: effects on energy partitioning and surface energy exchange in a high-Arctic tundra ecosystem, The Cryosphere, 10, 1395–1413, 2016.

Sturm, M., Holmgren, J., McFadden, J. P., Liston, G. E., Chapin III, F. S., and Racine, C. H.: Snow–shrub interactions in Arctic tundra: a hypothesis with climatic implications, Journal of Climate, 14, 336–344, 2001.

Tape, K., Sturm, M., and Racine, C.: The evidence for shrub expansion in Northern Alaska and the Pan-Arctic, Global Change Biology, 12, 686–702, doi:10.1111/j.1365-2486.2006.01128.x, http://dx.doi.org/10.1111/j.1365-2486.2006.01128.x, 2006.

Tucker, C. J., Slayback, D. A., Pinzon, J. E., Los, S. O., Myneni, R. B., and Taylor, M. G.: Higher northern latitude normalized difference vegetation index and growing season trends from 1982 to 1999, International Journal of Biometeorology, 45, 184–190, doi:10.1007/s00484-001-0109-8, http://dx.doi.org/10.1007/s00484-001-0109-8, 2001.

Uchida, M., Nakatsubo, T., Kanda, H., and Koizumi, H.: Estimation of the annual primary production of the lichen Cetrariella delisei in a glacier foreland in the High Arctic, Ny-Ålesund, Svalbard, Polar Research, 25, 39–49, 2006.

Uchida, M., Kishimoto, A., Muraoka, H., Nakatsubo, T., Kanda, H., and Koizumi, H.: Seasonal shift in factors controlling net ecosystem production in a high Arctic terrestrial ecosystem, Journal of plant research, 123, 79, 2009.

US Geological Survey, .: HYDRO1k Elevation derivative database, US Geological Survey Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota, https://lta.cr.usgs.gov/HYDRO1K, 2000.

van der Molen, M. K., van Huissteden, J., Parmentier, F. J. W., Petrescu, A. M. R., Dolman, A. J., Maximov, T. C., Kononov, A. V. ., Karsanaev, S. V., and Suzdalov, D. A.: The growing season greenhouse gas balance of a continental tundra site in the Indigirka lowlands, NE Siberia, Biogeosciences, 4, 985–1003, doi:10.5194/bg-4-985-2007, http://www.biogeosciences.net/4/985/2007/, 2007.

Van Huissteden, J., Maximov, T., and Dolman, A.: High methane flux from an arctic floodplain (Indigirka lowlands, eastern Siberia), Journal of Geophysical Research: Biogeosciences, 110, 2005.

Van Wijk, M. T., Williams, M., Laundre, J., and Shaver, G.: Interannual variability of plant phenology in tussock tundra: modelling interactions of plant productivity, plant phenology, snowmelt and soil thaw, Global Change Biology, 9, 743–758, doi:10.1046/j.1365-2486.2003.00625.x, http://dx.doi.org/10.1046/j.1365-2486.2003.00625.x, 2003.

Wang, T., Ottlé, C., Boone, A., Ciais, P., Brun, E., Morin, S., Krinner, G., Piao, S., and Peng, S.: Evaluation of an improved intermediate complexity snow scheme in the ORCHIDEE land surface model, Journal of Geophysical Research: Atmospheres, 118, 6064–6079, doi:10.1002/jgrd.50395, http://dx.doi.org/10.1002/jgrd.50395, 2013.

Weedon, G. P.: Readme file for the 'WFDEI' dataset, http://www.eu-watch.org/gfx_content/documents/README-WFDEI.pdf, 2013.

Weedon, G. P., Gomes, S., Viterbo, P., Shuttleworth, W. J., Blyth, E., Österle, H., Adam, J. C., Bellouin, N., Boucher, O., and Best, M.: Creation of the WATCH forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century, Journal of Hydrometeorology, 12, 823–847, doi:doi:10.1175/2011JHM1369.1, 2011.

Westermann, S., Lüers, J., Langer, M., Piel, K., and Boike, J.: The annual surface energy budget of a high-arctic permafrost site on Svalbard, Norway, The Cryosphere, 3, 245, 2009.

Westermann, S., Wollschläger, U., and Boike, J.: Monitoring of active layer dynamics at a permafrost site on Svalbard using multi-channel ground-penetrating radar, The Cryosphere, 4, 475–487, doi:10.5194/tc-4-475-2010, http://www.the-cryosphere.net/4/475/2010/, 2010.

Westermann, S., Boike, J., Langer, M., Schuler, T., and Etzelmüller, B.: Modeling the impact of wintertime rain events on the thermal regime of permafrost, The Cryosphere, 5, 1697–1736, 2011.

Westermann, S., Elberling, B., Højlund Pedersen, S., Stendel, M., Hansen, B. U., and Liston, G. E.: Future permafrost conditions along environmental gradients in Zackenberg, Greenland, The Cryosphere, 9, 719–735, 2015.

Wieder, W. R., Bonan, G. B., and Allison, S. D.: Global soil carbon projections are improved by modelling microbial processes, Nature Climate Change, 3, 909–912, 2013.

Wieder, W. R., Cleveland, C. C., Smith, W. K., and Todd-Brown, K.: Future productivity and carbon storage limited by terrestrial nutrient availability, Nature Geoscience, 8, 441–444, 2015.

Xenakis, G. and Williams, M.: Comparing microbial and chemical kinetics for modelling soil organic carbon decomposition using the DecoChem v1.0 and DecoBio v1.0 models, Geoscientific Model Development, 7, 1519–1533, doi:10.5194/gmd-7-1519-2014, http://www.geosci-model-dev.net/7/1519/2014/, 2014.

Yang, D., Kane, D., Zhang, Z., Legates, D., and Goodison, B.: Bias corrections of long-term (1973–2004) daily precipitation data over the northern regions, Geophysical Research Letters, 32, n/a–n/a, doi:10.1029/2005GL024057, http://dx.doi.org/10.1029/2005GL024057, l19501, 2005.

Yu, Z., Loisel, J., Brosseau, D. P., Beilman, D. W., and Hunt, S. J.: Global peatland dynamics since the Last Glacial Maximum, Geophysical Research Letters, 37, n/a–n/a, doi:10.1029/2010GL043584, http://dx.doi.org/10.1029/2010GL043584, l13402, 2010.

Yuan, W., Liu, S., Dong, W., Liang, S., Zhao, S., Chen, J., Xu, W., Li, X., Barr, A., Black, T. A., et al.: Differentiating moss from higher plants is critical in studying the carbon cycle of the boreal biome, Nature communications, 5, 2014.

Zhu, D., Peng, S., Ciais, P., Zech, R., Krinner, G., Zimov, S., and Grosse, G.: Simulating soil organic carbon in yedoma deposits during the Last Glacial Maximum in a land surface model, Geophysical Research Letters, 43, 5133–5142, 2016.

Zimov, S. A., Schuur, E. A., and Chapin III, F. S.: Permafrost and the global carbon budget, Science(Washington), 312, 1612–1613, 2006.

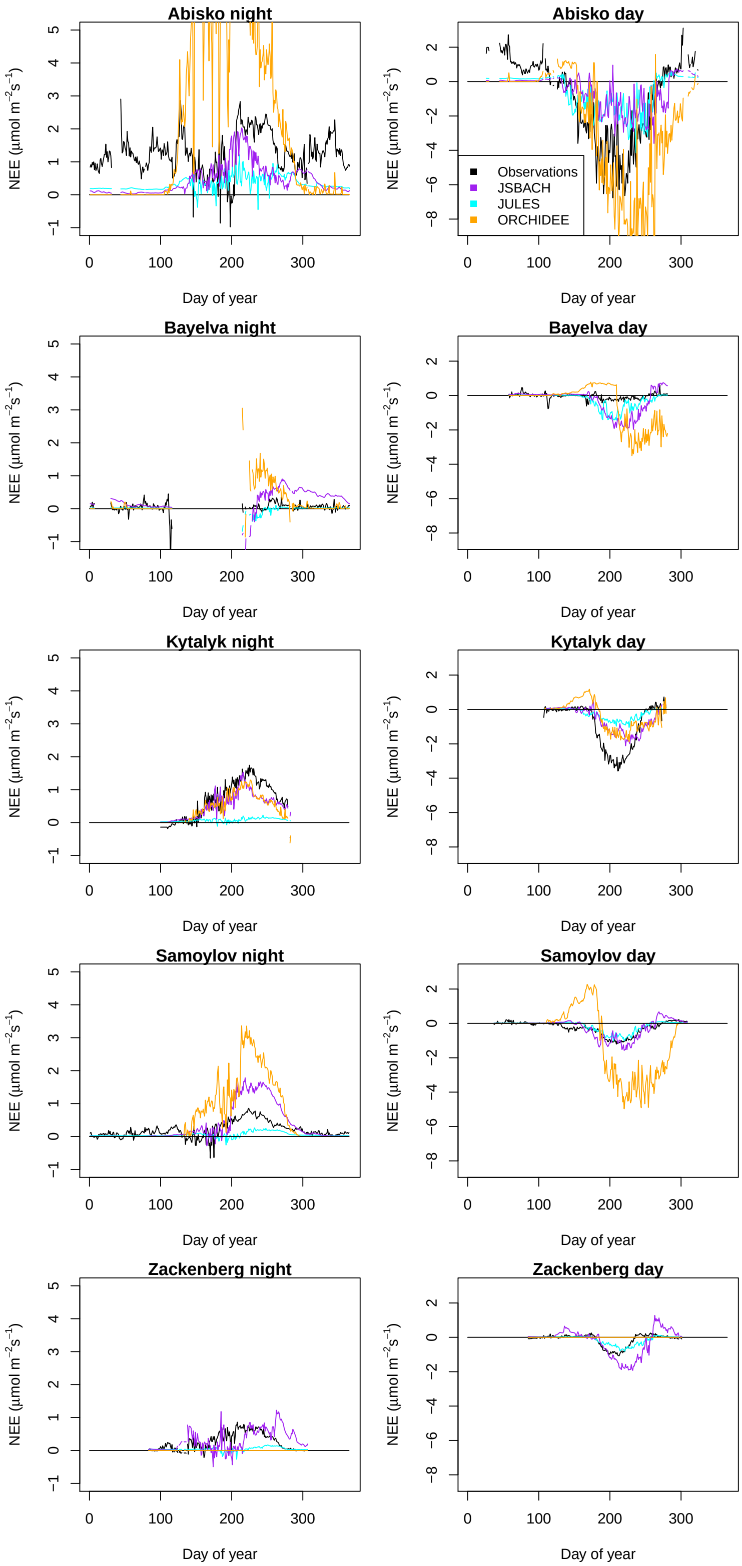


Figure 7. Mean annual cycles of CO_2 fluxes for all sites, observations and models. Left: nighttime flux; Right: daytime flux (corresponding to incoming shortwave radiation >20 Wm^{-2}). See supplementary Table S1 for years used at each site.

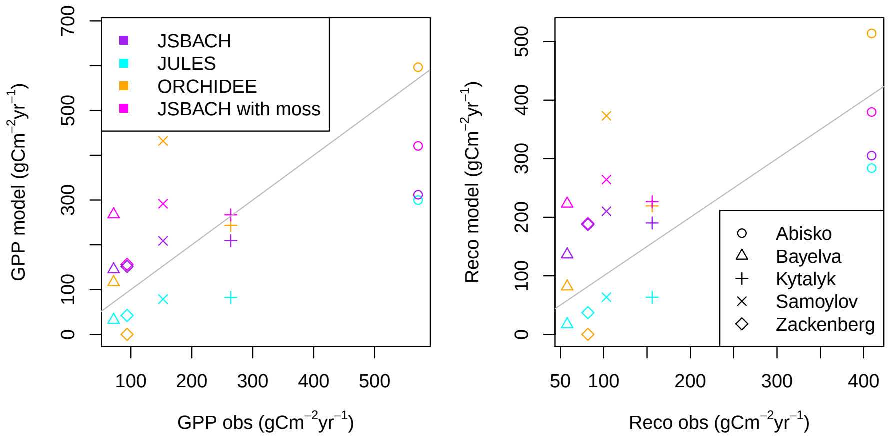


Figure 8. Mean annual GPP (gross primary productivity) and Reco (ecosystem respiration) from the models, plotted against the observation-derived values for the same time periods. See supplementary Table S1 for years used at each site.

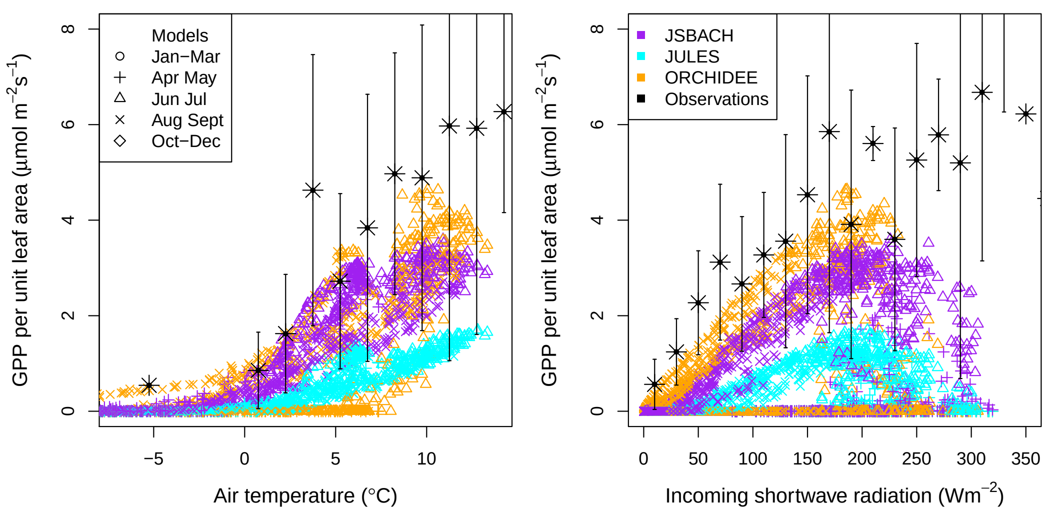


Figure 9. Relationship of ‘normalised’ GPP (GPP per m^2 of leaf) to air temperature and incoming solar radiation. All models and sites are shown, plus observationally-derived values using GPP estimated from eddy covariance data and LAI from MODIS (MODIS15A2), see Section 2.1.6.

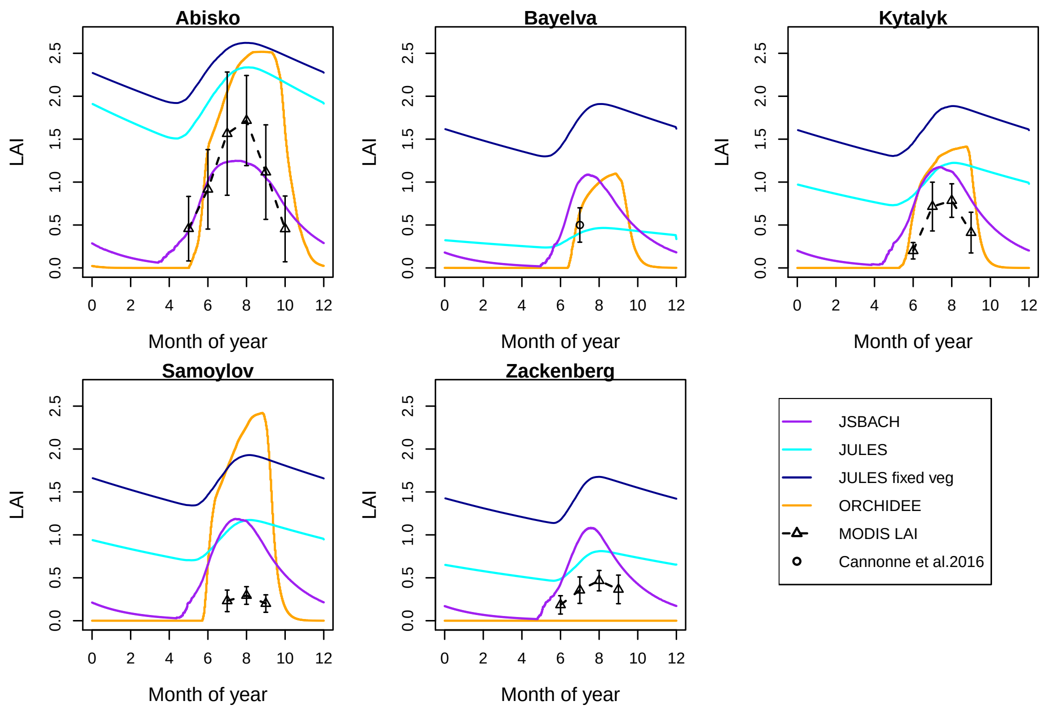


Figure 10. Mean annual cycles of LAI (leaf area index), for each site. ‘Observed’ values are from MODIS LAI product (MODIS15A2), except Bayelva which is from Cannone et al. (2016).

Sensitivity to snow

Representativeness of snow depths

In flat, open tundra landscapes, the snow is heavily affected by wind blowing, with the consequence that snow depth does not correspond directly to precipitation, and therefore using direct snowfall measurements is not possible in these landscapes. This scenario particularly applies to Kytalyk, Samoylov, and the Abisko mire. On a large scale, the snow can be quite even distributed due to the flat landscape (e.g. Blok et al., 2010, Table 2), but the microtopography at these sites (e.g. ice wedge polygons, palsas) leads to small-scale variability in snow depths. For example at Samoylov the depressed polygon centers have much deeper snow than the elevated rims (Boike et al., 2013). Thus a single point measurement of snow depth may not be representative of the whole flux tower footprint. At Abisko, however, several locations on the mire are averaged to give a representative sample.

At Bayelva and Zackenberg, the landscape is more mountainous, and there is more variation in snow depth around the area due to the topography of the land (and consequent differences in vegetation). At Zackenberg the snow is measured on transects across different vegetation types and the values range from snow-free to more than 1m of snow at a single time. However, the flux tower is situated in a fairly homogenous *cassiope* heath where snow surveys show the typical standard deviation of snow depth any one point in time is around +/-12cm (ZEROCALM1, https://data.g-e-m.dk/, average depth around 50cm). For this site, the point observation appears to be representative of the flux tower footprint. At Bayelva the snow depth varies by around +/-50% within the vicinity of the flux tower (Gisnås et al., 2014), and our point observation falls a little higher than the typical values for maximum snow height.

Even for sites where a point measurement of snow depth is representative of the flux tower area, the snowfall timeseries is derived using an assumed density and could be better parameterised using snow density measurements.

Sensitivity study

To investigate the impact of the variability and uncertainty in snow depth, we performed a sensitivity study. The observations suggest that increasing and decreasing the snow depth by 50% from the model simulated values would capture the range of observed snow depths in each of these landscapes. Since the snow depth is dynamically simulated rather than input to the models, we approximated the change in snow depth by increasing and decreasing the snowfall forcing by 50%. Two of the models (JSBACH and JULES) were then re-run (including spin-up) in these two different configurations. Snow depth in these simulations now spans a range that includes the point observations (Figure S5).

As expected, increased snow depth leads to an overall warming of the soil for every site, and reduced snow depth leads to a cooling (Figure S6). However, most of the change happens in winter, where it will have less impact on the carbon cycle since the vegetation and soil decomposition processes take place mainly in summer (JJA) (Figure S6).

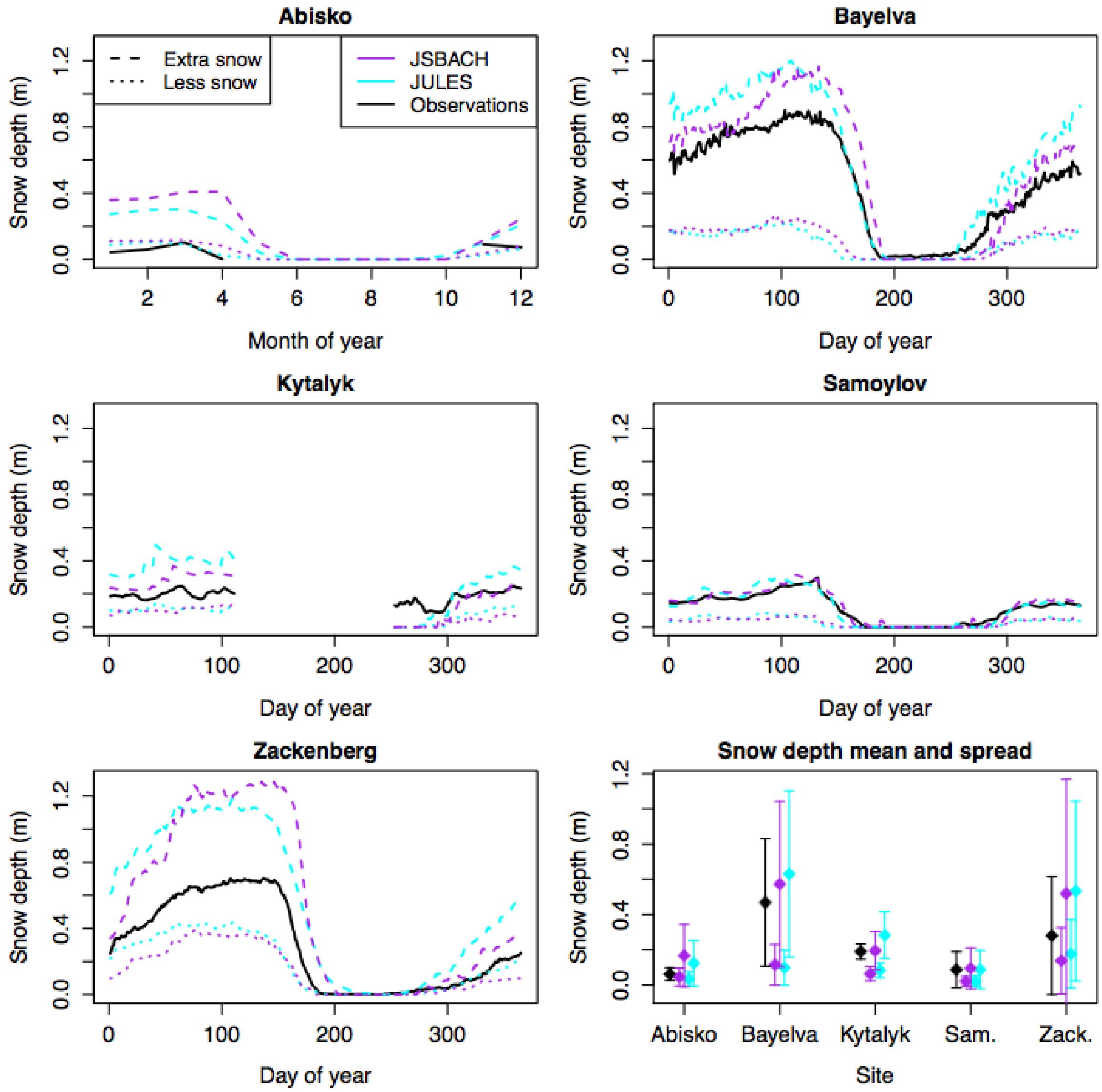


Figure S5 | Mean annual cycles of snow depth (as in Fig. 1 in main manuscript) showing simulations with increased and reduced snowfall in JSBACH and JULES.

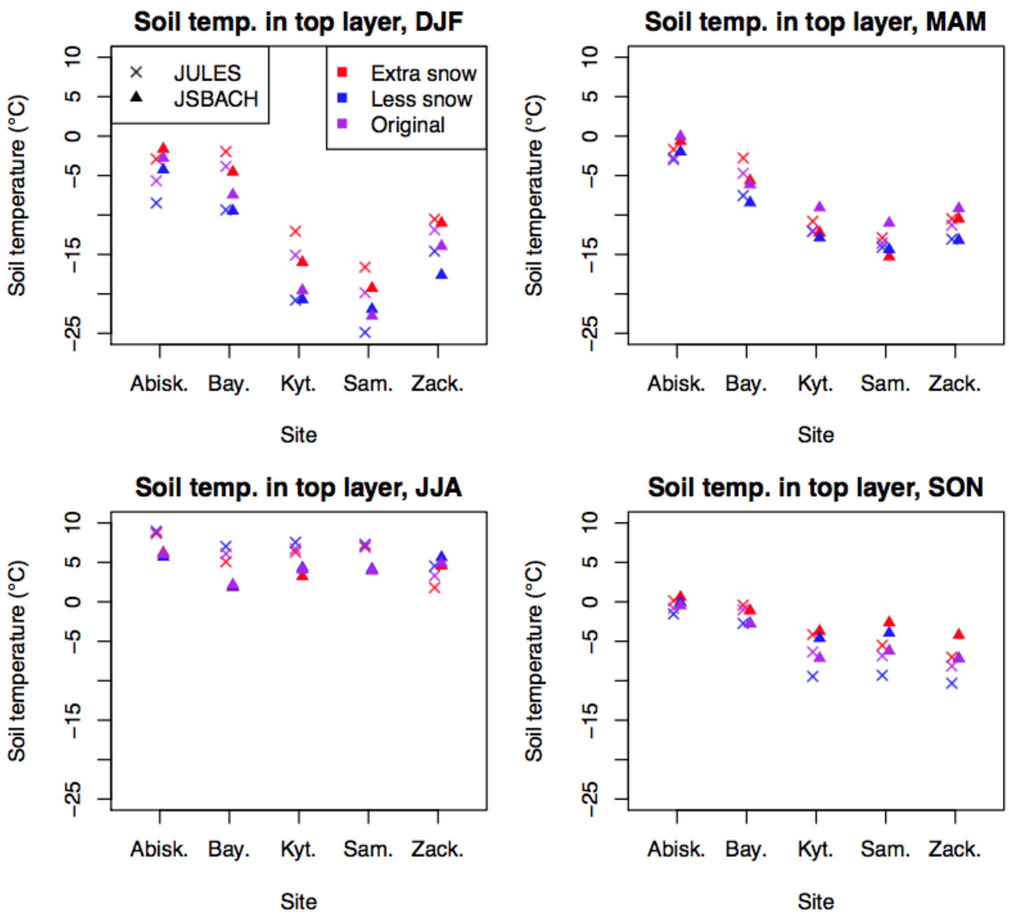


Figure S6 | Mean soil temperature in different seaons, showing simulations with increased and reduced snow for JULES and JSBACH. (DJF=December, January, February. MAM=March, April, May. JJA=June, July, August. SON = September, October, November.)

Vegetation growth is not directly impacted by snow or soil temperature changes in these models. However, the change in winter snowfall also leads to changes in soil moisture during summer, which does affect vegetation growth. An increase in snow should lead to an increase in water infiltration into the soil in spring and thus an increase in the available soil moisture. In JULES, however, for two of the sites (Zackenberg and Bayelva) the opposite effect is seen, where increased snow depth leads to less soil moisture in summer, and vice versa (Figure S7). In JULES, the changes in soil moisture are reflected in the GPP, ecosystem respiration (Reco) and vegetated fraction, which all increase with higher soil moisture and reduce with lower soil moisture (Figure S7). At many of the sites these are significant changes (although they still leave the model with low values of GPP/Reco compared to observed fluxes). The impact of any change in GPP is amplified by the resulting changes in vegetation fraction. In JSBACH, however, the changes in soil moisture, GPP and Reco are not significant (Figure S7).

Soil carbon stocks are impacted directly by the soil thermal state (as well as soil moisture, and inputs from vegetation). For JSBACH, while the vegetation fluxes do not show any noticeable sensitivity to snowfall (Figure S7), the soil carbon has a small but consistent trend towards lower soil carbon in the simulations with increased snow (Figure S8), which – since the other influencing variables have not significantly changed – is most likely due to consistently higher soil temperatures when more snow is present. For JULES, however, any changes in decomposition due to soil temperature are obscured by larger differences of vegetation inputs, particularly for Kytalyk and Samoylov sites (Figure S8), where the vegetation fractions are very different during spinup for the different sensitivity tests, and thus the rate of soil carbon accumulation changes significantly.

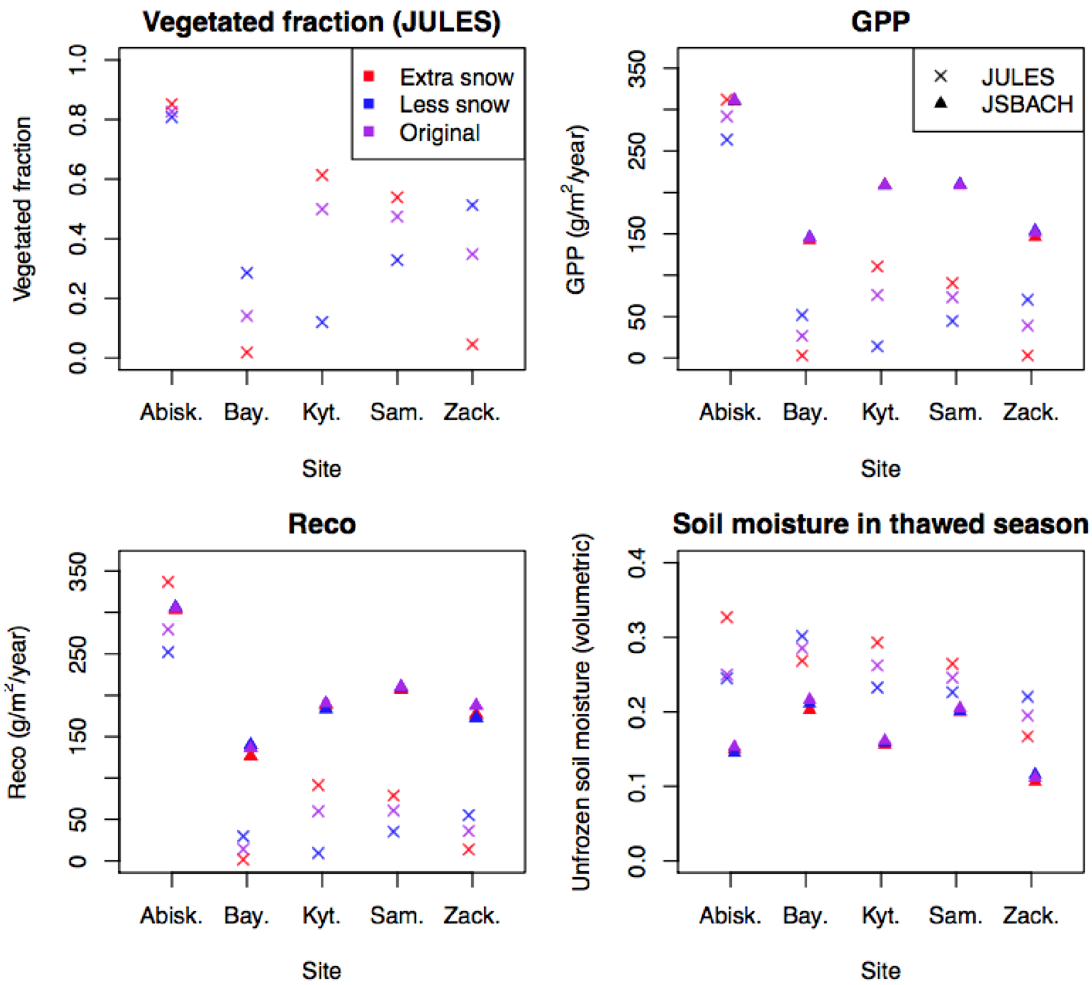


Figure S7 | Impacts of increased/reduced snowfall on soil moisture and carbon-cycle related variables (GPP, ecosystem respiration, and vegetated fraction), in JSBACH and JULES.

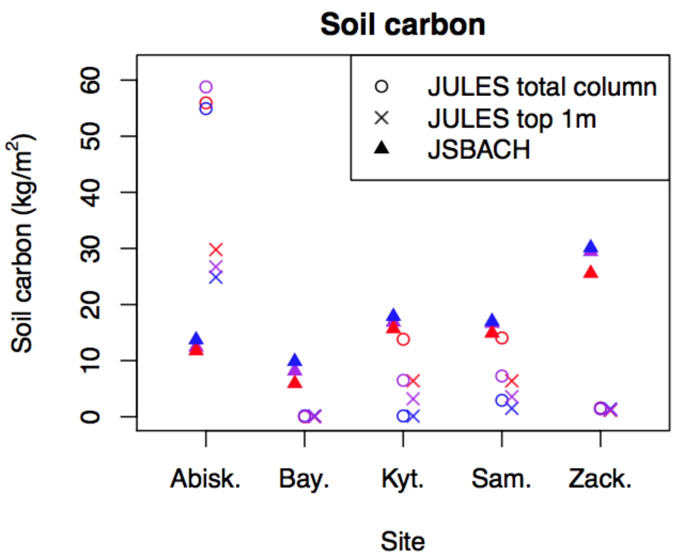


Figure S8 | Impact of increased/reduced snowfall on soil carbon stocks in JSBACH and JULES.

Discussion

Our sensitivity study has shown a high sensitivity of surface soil temperature to a 50% change in snow depth of up to 5ºC or more, seasonally. This is in line with observations (Gisnås et al., 2014). Soil carbon decomposition is sensitive to these soil temperature changes, resulting in lower carbon stores for the warmer simulations in JSBACH (Fig. S8), which is in line with studies such as Lund et al. (2012) which showed that snow affected the carbon budget at Zackenberg by warming the soil and increasing soil respiration. However, the impact of snow on soil moisture is not in line with observed behaviour: in general, more snow should lead to increased soil moisture availability in summer (see for example Litaor et al., 2008). However, in JULES for two of the sites, the summer soil moisture is reduced with additional snowfall, and in JSBACH there are no significant changes. This supports the conclusion that more work is needed on the hydrology schemes in these models. Furthermore, the models are missing some snow-vegetation interactions such as preventing vegetation growth when covered by snow, or protection from damage in winter.

It is also important for the models to better represent the profile of snow thermal conductivity: for example the models do not simulate the low-conductivity 'depth-hoar' layer that can form at the base of the snowpack (Domine et al., 2016). For this, monitoring of snow temperature at different heights can be valuable to improve the models (Barrere et al., 2017). It is also useful to compare snow density in models and observations. For example, recent work shows that including wind compaction is essential to capture high snow density at Samoylov (Gouttevin et al., 2017), and indeed our models show a snow density closer to the 'default' model in Gouttevin et al. (2017), which is too low due to omission of wind effects.

In large-scale modelling, it is certainly important to represent variability in snow depth, which is only coarsely included in land surface models in most cases (e.g. snow depth varies only between surface tiles (Essery et al., 2003)). For recent developments towards this, see for example Gisnås et al. (2014).

References

Barrere, M., *et al*.: Evaluating the performance of coupled snow-soil models in SURFEXv8 to simulate the permafrost thermal regime at a high Arctic site, *Geosci. Model Dev. Discuss.*, 2017, 1-38, 2017.

Blok, D., *et al*.: Shrub expansion may reduce summer permafrost thaw in Siberian tundra, *Global Change Biol.*, **16**(4), 1296–1305, doi:10.1111/j.1365-2486.2009.02110.x, 2010.

Boike, J., *et al.*: Baseline characteristics of climate, permafrost and land cover from a new permafrost observatory in the Lena River Delta, Siberia (1998–2011), *Biogeosciences*, **10**, 2105-2128, https://doi.org/10.5194/bg-10-2105-2013, 2013.

Domine, F., *et al.*: Seasonal evolution of the effective thermal conductivity of the snow and the soil in high Arctic herb tundra at Bylot Island, Canada, *The Cryosphere*, **10**, 2573–2588, doi:10.5194/tc-10-2573-2016, 2016.

Essery, R.L., *et al*.: Explicit Representation of Subgrid Heterogeneity in a GCM Land Surface Scheme. *J. Hydrometeor.*, **4**, 530-543, 2003.

Gisnås, K., *et al*.: A statistical approach to represent small-scale variability of permafrost temperatures due to snow cover, *The Cryosphere*, **8**, 2063-2074, https://doi.org/10.5194/tc-8-2063-2014, 2014.

Gouttevin, *et al.*: Observation and modeling of snow at a polygonal tundra permafrost site: thermal implications and spatial variability. Submitted to *The Cryosphere*, 2017.

Litaor, M. I., *et al*.: Topographic controls on snow distribution, soil moisture, and species diversity of herbaceous alpine vegetation, Niwot Ridge, Colorado, *J. Geophys. Res.*, 113, G02008, doi:10.1029/2007JG000419, 2008.

Lund, M., *et al*.: Trends in CO_2 exchange in a high Arctic tundra heath, 2000–2010, *J. Geophys. Res.*, **117**, G02001, doi:10.1029/2011JG001901, 2012.