Response to reviewers of “The GRENE-TEA Model Intercomparison Project (GTMIP): Overview and experiment protocol for Stage 1,” by Miyazaki et al.

We greatly appreciate Referee #1 for his/her positive evaluation on the framework, targets and analysis plan of the intercomparison project, and for helpful and constructive comments and suggestions on the discussion paper.

Overall, we have substantially reorganized and rephrased the manuscript; based on the valuable comments from the two referees to clearly present that this is such a paper as for ”model intercomparison descriptions, including experimental details and project protocols” (GMD’s Aims and scope). Below, referee comments are shown in red and italic, and our specific replies follow.

Referee #1

The paper describe the protocol and very preliminary results for the stage 1 of the GRENE-TEA model inter-comparison project. The proposed inter-comparison project is very interesting as it aims to compare very different kinds of models in their ability to simulate both biophysical and biogeochemicals processes of the pan-artic region. These region is obviously of first importance since it should experience the highest temperature change in the future. Because of the permafrost, there is a risk of large feedback with soil and soil carbon and then it is a very vulnerable region. Moreover the models have not been extensively evaluated for these regions. So this is and interesting subject and I am sure that project could lead to very interesting results.

However the project is still at a very early stage. Then my main concern about the paper in it present form is that there is almost no results presented. Then I really not understand why the authors want to publish a paper at this early stage of the project and not waiting the end of the stage1 to present a more complete analysis of the results ? I would eventually understand if the experiment protocol leads to development of specific new tools. But this is not the case *here where the protocol is relatively standard for such kind of experiment. The paper gives a promising analysis plan, looking to the cause of differences between models, studies at different temporal time scale and conducting several sensitivity tests. Then part dedicated to presentation of results is very frustrating as it is less than one page and stay very descriptive without any real discussion about results. For all these reasons I think that the paper cannot be published in its present form and should be resubmitted with a complete analysis of the stage 1 results when, I am sure it will be a very interesting and useful contribution for the modelling community.*

We thank Referee #1 for his/her very positive and encouraging evaluation on this model intercomparison project. We think, however, the publication of the protocol as a separate paper is grounded and reasonable for the following reasons:

1. As stated more clearly in the revised new section “3. Analysis plan and exemplary results” (an amalgamation and expansion of the previous Sections 2.4 and 3), the analyses and the consequent resulting papers will vary in topics, each of which will be authored by a different group of researchers, thereby this protocol paper acts as a guiding paper.
2. This paper also provides an overview of the entire GTMIP activity (including both Stages, 1 and 2), which include the frameworks of the model intercomparison *per se*, as well as the collaborations among modelling and empirical communities in Japan that assisted the activity (cf. new Fig. 1). Creation of driving data (for spin-up and the experiment), and the boundary conditions on ecosystem activities is already a product of such collaborations, and described in the companion paper (Sueyoshi et al., accepted for publication as a discussion paper in Earth System Science Data Discussion (ESSDD)). (new Section 1.2 GRENE-Arctic project and GTMIP)
3. This model intercomparison is a unique project among previous MIPs in terms of its scope (ranging from physical to ecosystem processes) and geographical domain (the Arctic region) of the target (revised Section 1. Introduction).
4. Publication of the protocol is a part of interactions in the project, and a critical mean for recruiting new participants. Participation in this GTMIP stage 1 is voluntary and open to any interested parties (modellers, groups, and/or institutions). Actually, a new participant joined after the publication of the discussion paper.

Considering those points and issues, we have reorganized and rephrased the manuscript. Major revision has been done in Abstract, and Introduction, which are given below:

- Abstract

"As part of the terrestrial branch of the Japan-funded Arctic Climate Change Research Project (GRENE-TEA), which aims to clarify the role and function of the terrestrial Arctic in the climate system and assess the influence of its changes on a global scale, this model intercomparison project (GTMIP) is deliberatively designed to 1) enhance communication and understanding between the "minds and hands" (i.e., between the modelling and field scientists) and 2) assess the uncertainty and variations stemming from variability in model implementation/design and in model outputs using climatic and historical conditions in the Arctic terrestrial regions. This paper provides an overview of all GTMIP activity, and the experiment protocol of Stage 1, which is site simulations driven by statistically fitted data created using the GRENE-TEA site observations for the last three decades. The target metrics for the model evaluation cover key processes in both physics and biogeochemistry, including energy budgets, snow, permafrost, phenology, and carbon budgets. Exemplary results for distributions of four metrics (annual mean latent heat flux, annual maximum snow depth, gross primary production, and net ecosystem production), and for seasonal transitions are provided to give an outlook of the planned analysis that will delineate the inter-dependence among the key processes, and provide clues for improving model performance."

- 1. Introduction

"The pan-Arctic ecosystem is characterized by low mean temperatures, snow cover, and seasonal frozen ground or permafrost with a large carbon reservoir, covered by various biomes (plant types) ranging from deciduous and evergreen forests to tundra. The Arctic climate and ecosystem differ from the tropical and temperate counterparts primarily because it is a frozen world. Moreover, the terrestrial Arctic varies from area to area according to the location, glacial history, and climatic conditions. However, sites, networks, and opportunities for direct observations are still sparse relative to the warmer regions owing to physical and logistical limitations. To investigate the impact of climate change in this region, a number of studies using both analysis of observed data and numerical modelling have been carried out (e.g., Zhang et al., 2005; Brown and Robinson, 2011; Brutel-Vuilmet et al., 2013; Koven et al., 2011, 2013; Slater and Lawrence, 2013). Various numerical modelling schemes have been developed to treat physical and biogeochemical processes on and below the land surface. Some of these processes are site-specific or process-oriented, while others are implemented as components of atmosphere–ocean coupled global climate models (AOGCMs), or Earth system models (ESMs) to interact with the overlying atmosphere. Among these processes, snowpack, ground freezing/thawing, and carbon exchange are the most relevant and important processes in terrestrial process models (TPM) for investigating the climate and ecosystem of the pan-Arctic region.

1.1 GRENE-Arctic project and GTMIP

The GRENE-TEA model intercomparison project (GTMIP) was originally planned as part of the terrestrial research project of the GRENE Arctic Climate Change Research Project (GRENE-TEA) to achieve the following targets: a) to pass possible improvements regarding physical and biogeochemical processes for Arctic terrestrial modelling (excluding glaciers and ice sheets) in the existing AOGCM terrestrial schemes for the AOGCM research community, and b) to lay the foundations for the development of future-generation Arctic terrestrial models. The project, however, involves groups of researchers from different backgrounds/disciplines (e.g., physics/geophysics, glaciology, biogeochemistry, ecosystem, forestry) with a wide range of research methods (e.g., field observations, remote-sensing, numerical modelling), target domains (e.g., Northern Europe, Siberia, Alaska, Northern Canada) and scales (from site-level to Pan-Arctic). As is often the case, multi-disciplinary opportunities were limited, initially creating a considerable challenge for the project (Fig. 1a). Communications between groups (e.g., modelling and field studies, physical and ecosystem disciplines, process-oriented and large-scale modelling), if any, were inconclusive and sporadic. Observational practices and procedures (e.g., variables to measure, equipment to use, standard zero depth for ground measurements) were different among groups and disciplines, and lacked standardization. Although each individual group had the needs and intention to interact with other groups, the requisite collaboration could not be achieved. Opinions obtained in the early stages revealed hidden quests for possible collaborations for “observational data for driving and/or validating data”, “use of numerical models to test empirical hypothesis gained at the field”, “interpretation of observed phenomena”, and “optimization of observation network strategies.” As a result of this situation, the model intercomparison project was deliberately blueprinted to promote communication and understanding between modelling and empirical scientists, and among modellers: the GTMIP protocols and datasets are set to function as a hub for the groups involved in the project (Fig.

1b). It also aimed to enhance the standardization of observation practices among the GRENE-TEA observation sites, and to form a tight collaboration between the field and modelling communities, laying a cornerstone for creating the driving dataset (details of the Stage 1 driving data and their creation as a product of collaboration between modellers and field scientists are documented by Sueyoshi et al. [2015]).

1.2 Model intercomparison for the terrestrial Arctic

Since the 1990s, a number of model intercomparison projects (MIPs) have been carried out, focusing on the performance of TPMs, AOGCMs, and ESMs; examples include PILPS (Project for Intercomparison of Land-Surface Parameterization Schemes; Henderson-Sellers, 1993), SnowMIP (Snow Models Intercomparison Project; Etchevers et al. 2004; Essery et al. 2009), Potsdam NPP MIP (Potsdam Net Primary Production Model Intercomparison Project; Cramer et al., 1999), C4MIP (Coupled Climate–Carbon Cycle Model Intercomparison Project; Friedlingstein et al. 2006), CMIP5 (Coupled Model Intercomparison Project; Taylor et al. 2012), and MsTMIP (Multi-scale synthesis and Terrestrial Model Intercomparison Project; Huntzinger et al., 2013), to name a few.

For snow dynamics, SnowMIP2 showed a broad variety in the maximum snow accumulation values, particularly at warmer sites and in warmer winters, although the duration of snow cover was relatively well simulated (Essery et al., 2009). The same study also noted that the SnowMIP2 models tend to predict winter soil temperatures that are too low in cold sites and for sites with shallow snow, a discrepancy arguably caused by the remaining uncertainties in ecological and physical processes and the scarcity of winter process measurements for model development and testing in the boreal zone. The CMIP5 models simulated the snow cover extent for most of the Arctic region well, except for the southern realm of the seasonal snow cover area (Brutel-Vulmet et al., 2013). The poor performance of some of the TPMs in this region is due to an incorrect timing of the snow onset, and possibly by an incorrect representation of the annual maximum snow cover fraction (Brutel-Vulmet et al., 2013). For ground freezing/thawing processes, Koven et al. (2013) showed the current status of the performance of AOGCMs for permafrost processes based on CMIP5 experiments. There was large disagreement among modelled soil temperatures, which may have been due to the representation of the thermal connection between the air and the land surface and, in particular, its mediation by snow in winter. Vertical profiles of the mean and amplitude of modelled soil temperatures showed large variations, some of which can be attributed to differences in the physical properties of the modelled soils and coupling between energy and water transfer. This appears to be particularly relevant for the representation of organic layers.

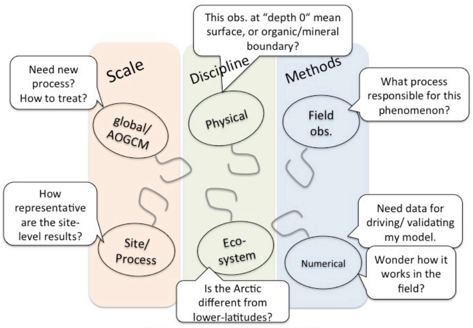
For the biogeochemical cycles, a number of studies based on MIPs have been carried out. The broad global distribution of net primary productivity (NPP) and the relationship of annual NPP to the major climatic variables coincide in most areas with differences among the 17 global terrestrial biogeochemical models that cannot be attributed to the fundamental modelling strategies (Cramer et al., 1999). The ESMs in CMIP5 use the climate and carbon cycle performance metrics, and they showed that the models correctly reproduced the main climatic variables controlling the spatial and temporal characteristics of the carbon cycle (Anav et al., 2013). However, several weaknesses were found in the modeling of the land carbon cycle: for example, the leaf area index is generally overestimated by models compared with remote sensing data (Anav et al., 2013); NPP and terrestrial carbon storage responses to CO2 increases greatly differs among models (Hajima et al., 2014); current ESMs displays large variations for the estimated soil carbon amounts, in particular for northern high latitudinal regions, and lack the capability to represent the potential degradation of frozen carbon in permafrost regions (Todd-Brown et al., 2014). The future projection by ESMs suggests that the carbon sink characteristic will increase in northern high latitudes, although there are some uncertainties, such as nutrient limitations in CO2 fertilization, the effect of soil moisture on decomposition rates, and mechanistic representations of permafrost (Qian et al., 2010; Ahlstrom et al., 2012; Arora et al., 2013). It should be noted that the reference observation data used for these evaluations are prone to uncertainties due to random and bias errors in the measurements themselves, sampling errors, and analysis error, especially for biogeochemical variables such as land gross primary productivity (GPP) (e.g., Anav et al., 2013; Piao et al., 2013). Based on the outcomes of these MIPs, TPMs have improved their performances.

At scales from a continental level (including those mentioned above) to a site level (model-observation comparisons; e.g., Zaehle et al., 2014), different MIPs have also been conducted, and generally study physical or ecosystem processes separately. PILPS (Henderson-Sellers et al., 1993) and a series of snow MIPs (Etchevers et al., 2004; Essery et al., 2009) are well-known MIPs for physical processes, targeting hydrology and snow dynamics. Recently, an MIP for tundra sites has been conducted, but its focus is limited to soil thermal dynamics

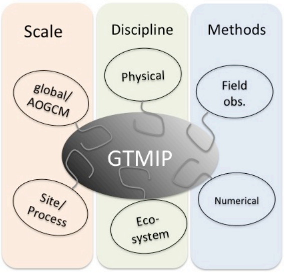
(Ekici et al., 2014). In turn, ecosystem MIPs on continental scales have two predecessors: i.e., the North American Carbon Program Site Synthesis (Schwalm et al., 2010) and CarboEastAsia-MIP (Ichii et al., 2013). Although both MIPs employ multiple terrestrial biosphere models to different eddy-covariance measurement sites (Schwalm et al. (2010) with 22 models for 44 sites in North America; Ichii et al. (2013) with 8 models for 26 sites in East Asia), boreal and Arctic sites were not the major targets. In other studies targeting specific eco-climatic regions, the Arctic was again not the main domain: Jung et al. (2007) assessed GPPs for Europe, and Ichii et al. (2010) for Japan. Rawlins et al. (2015) assessed carbon budget differences among several GCM-compatible models in northern Eurasia, with little examination of the physical processes. In other regions than the Arctic, there have been cross-sectional evaluations of physical and ecosystem processes, such as Morales et al. (2005), evaluating carbon and water fluxes in Europe, and de Gonçalves et al. (2013), the LBA-Data Model Intercomparison Project (LBA-DMIP), analysing water and carbon fluxes in the Amazon.

The GTMIP consists of two stages (Fig. 2): one dimensional, historical GRENE-TEA site evaluations for examining the model's behaviour and its uncertainty (Stage 1), and circumpolar evaluations using projected climate change data from GCM outputs (Stage 2). Hereafter, we describe the Stage 1 protocol. This stage aims to evaluate the physical and biogeochemical TPMs through three-decade site simulations driven and validated by the GRENE-TEA site-derived data. It calls for broader participation in the activity from a wider community to assure robust assessments for model-derived uncertainty, and to efficiently investigate the terrestrial system response to climate variability considering the diversity of the pan-Arctic sites. Thus, the scope and geographical domain of GTMIP Stage 1 is unique in its target of the Arctic region, including both taiga and tundra, and in its evaluations of the behaviour of the energy-snow-soil-vegetation subsystem, employing a wide range of models from physical land surface schemes to terrestrial ecosystems. "

- New Figures 1 and 2:



"Pirates" before GTMIP



At the "Round Table"

Figure 1. "Pirates of the Arctic" sit together at thea Round Table

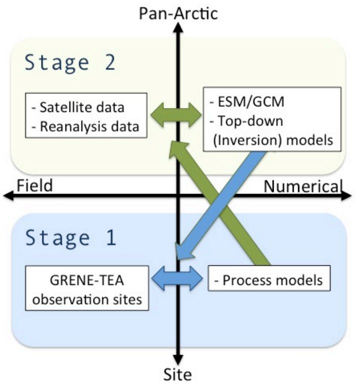


Figure 2. Schematic diagrams for stage 1 and 2 of GTMIP

Since this paper is intended to be a model experiment description paper, the "Preliminary results" section in the original manuscript was meant to provide "sample model output" as "descriptions/figures of model results to give an overview of the project." In order to addressed the concern of the referee that the "*part dedicated to presentation of results is very frustrating as it is less than one page and stay very descriptive without any real discussion about results*" we have merged the "2.4 Analysis plan" section to the new Section "3. Analysis plan and exemplary results", and showed more clearly the descriptions of the model results with sample figures for topical analyses (Figures 5-8), and cross-sectional analyses (Figure 9; seasonal transitions).

- 3. Analysis plan and exemplary results

“This section presents the analysis plan for GTMIP Stage 1 and sample outputs based on already submitted materials. To answer the key questions for the target processes proposed in Sect. 2.1, we plan to analyze the model output by describing the model–model and model–observation differences, discerning the cause of these differences, and investigating parameter sensitivity. The outputs of multiple models will be compared in terms of the metrics shown in Table 3. These metrics are divided into five categories (i.e., energy and water budget, snowpack, phenology, subsurface hydrological and thermal states, and carbon budget). For terrestrial climate simulations on the decadal scale, the most important outputs are the latent heat flux (energy and water budget) and the net ecosystem exchange (carbon budget). The latent heat flux (evapotranspiration) is the essential driver of precipitation inland at high latitudes owing to high rates of recycling (e.g., Dirmeyer et al., 2009; Saito et al. 2006). Net ecosystem exchange (NEE) plays a fundamental role in determining global CO_2 concentrations by determining whether a site forms a carbon source or sink (e.g. Abramowitz et al., 2008; Mcguire et al., 2012). NEE represents the net land–atmosphere CO_2 flux, and a positive NEE represents net loss of CO_2 from the land to the atmosphere (i.e., carbon source; Mcguire et al., 2012). Although NEE is commonly used for tower flux observations and some TPMs, the net ecosystem production (NEP) is used in GTMIP for both the observed and simulated values because it is more widely used in non-biogeochemical communities. A positive (negative) value of NEP represents a carbon sink (source).

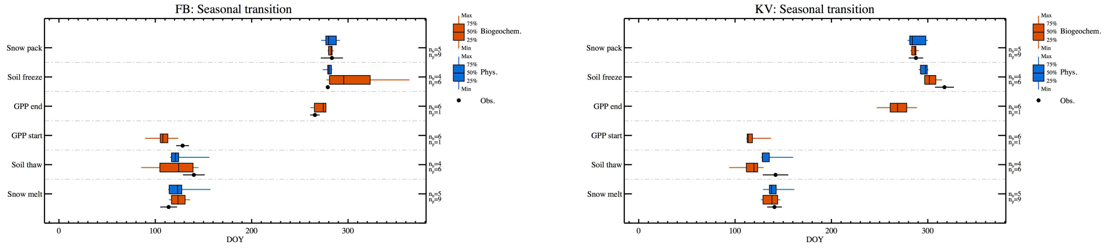
Analyses will be organized and conducted in the following manner. Topical analyses, constituting major subsets of the project outcomes, will evaluate characteristics of model performances and their inter-site variations within each of the above five categories, while cross-sectional analyses between categories will explore the functionality and strength of interactions between processes. These analyses will be utilized for mining crucial processes to improve the site-level TPMs as well as large-scale GCM/ESM components.

First, the focus will be on model output variability for both the inter-annual and the inter-decadal time scales, based on the output time series over more than 30 years. Inter-site differences will also be evaluated for the four GRENE-TEA sites in the Arctic region, each of which has distinct characteristics. The vegetation type for three of the four sites is forest (two evergreen conifer: FB and KV; one deciduous conifer: YK) and the remaining site is tundra (TK). Three sites (FB, TK, and YK) are in the permafrost region, while KV is underlain by seasonally frozen ground. Figures 5–8 show statistical summary comparisons of the model outputs by site (the land cover and soil type parameters used for the simulations are shown in Table 2), expressing inter-model variations for physical and biogeochemical models using box plots for four variables of the metrics mentioned above: the annual mean latent heat flux (Qle_total_an), the annual maximum snow depth (SnowDepth_max), the annual gross primary production (GPP_an), and the annual net ecosystem production (NEP_an), respectively. When observed values were available (i.e., latent heat flux for FB for 2011–2013 and YK for 1998, 2001, 2003, 2004, 2007, and 2008), they are shown by black dots.

Second, the cause or attributes of the differences among models, or between models and observations, will be explored by employing statistical evaluations such as multivariate analyses and time series analyses on the metrics and individual eco-climate variables. This will improve understanding of the interrelation between the incorporated processes in each model. Figure 9 shows an exemplary comparison of a seasonal transition in the snow-permafrost-vegetation sub-system, expressed similarly by box plots. The figure summarizes the average dates for (from bottom to top) the completion of snow melt, the thawing of the top soil layer, the start and end of greening, the freezing of the top soil layer, and the start of seasonal snow accumulation. A comparison of the timings of these events over years and sites will illustrate individual models' characteristic behaviour in seasonal transitions, and their strength regarding process interactions, in combination with ordinary multivariate analysis techniques.

Finally, sensitivity tests for the model parameters are planned to quantify the effect of parameter sensitivity on the models' reproducibility. ”

- New Figure 9



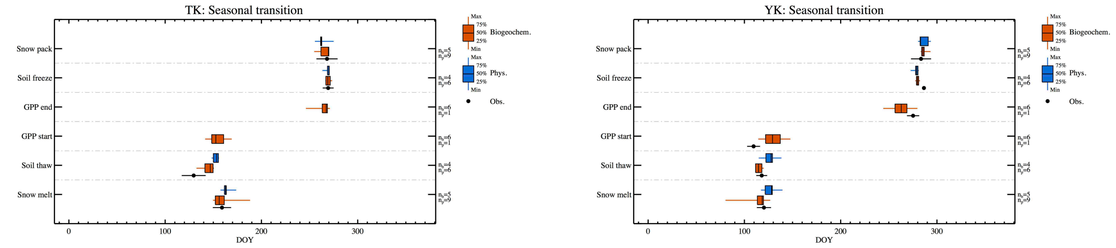


Figure 9. Example of seasonal transitions in ground temperature, snow, and vegetation among models.

Response to reviewers of “The GRENE-TEA Model Intercomparison Project (GTMIP): Overview and experiment protocol for Stage 1,” by Miyazaki et al.

We greatly appreciate Referee #2 for his/her positive evaluation on the framework, targets and analysis plan of GTMIP, and for helpful and constructive comments and suggestions on the discussion paper.

Overall, we have substantially reorganized and rephrased the manuscript, based on the valuable comments from the two referees, to clearly present that this is such a paper as for ”model intercomparison descriptions, including experimental details and project protocols” (GMD’s Aims and scope). Below, referee comments are shown in red and italic, and our specific replies follow.

Referee #2

General comment:

This paper suffers from a confusive presentation. What do the authors want to present? A protocol or results?

After receiving similar comments from both the referees, we reconsidered the organization and presentation of the discussion paper. We have substantially revised the manuscript (including Abstract), following the referees’ comments and advice, to make our intention of a “protocol” paper clear.

In the abstract it sais this paper presents an experiment protocol, but alo preliminary results are mentioned. The main body of the paper focusses on the experiments (setup, datasets, plan, output) and on the pre-liminary results, which gives the reader the impression that the main focus is not the protocol, but the experimentation results. This impression is additionaly by the authors phrasing. For example, page 6, line 20: this paper focusses on stage 1 of the project, which evaluates the TPMs.... etc.". Another example are the scientific questions at page 7, lines 10-14. Lines 13-14 are specificially important, because that is what is exactly lacking in the paper (which processes are responsible for model performance), if the focus is *on results rather then on protocol. A last example is the language used: page 10, line 25 "we will examine", page 11, line 8 "we will conduct". This is the kind of phrasing which is being used in a result-oriented paper. However, as a result-oriented paper this manuscript is insufficient because all results are preliminary, and the results are not discussed. Based on this I would reject the paper.*

We have reorganized and rephrased to clarify that our focus is to describe the overview of the GTMIP activity and the protocol (that is, framework, targets, and analysis plan) of its Stage 1. We have merged the Analysis plan section and the preliminary results section into a new “Analysis plan and exemplary results” to clearly show “descriptions/figures of model results to give an overview of the project”.

The summary (4) is however quite clear, this paper wants to present a protocol, and the preliminary results are an example.

My advice is to rewrite the paper, and make the audience more clear what your intentions are, right from the start of the manuscript.

We have accordingly revised the manuscript to clarify the intension of the project, and the objective of the manuscript. Major revisions are as follows.

- Abstract

“As part of the terrestrial branch of the Japan-funded Arctic Climate Change Research Project (GRENE-TEA), which aims to clarify the role and function of the terrestrial Arctic in the climate system and assess the influence of its changes on a global scale, this model intercomparison project (GTMIP) is deliberatively designed to 1) enhance communication and understanding between the "minds and hands" (i.e., between the modelling and field scientists) and 2) assess the uncertainty and variations stemming from variability in model implementation/design and in model outputs using climatic and historical conditions in the Arctic terrestrial regions. This paper provides an overview of all GTMIP activity, and the experiment protocol of Stage 1, which is site simulations driven by statistically fitted data created using the GRENE-TEA site observations for the last three decades. The target metrics for the model evaluation cover key processes in both physics and biogeochemistry, including energy budgets, snow, permafrost, phenology, and carbon budgets. Exemplary results for distributions of four metrics (annual mean latent heat flux, annual maximum snow depth, gross primary production, and net ecosystem production), and for seasonal transitions are provided to give an outlook of the planned analysis that will delineate the inter-dependence among the key processes, and provide clues for improving model performance."

- 1. Introduction

"The pan-Arctic ecosystem is characterized by low mean temperatures, snow cover, and seasonal frozen ground or permafrost with a large carbon reservoir, covered by various biomes (plant types) ranging from deciduous and evergreen forests to tundra. The Arctic climate and ecosystem differ from the tropical and temperate counterparts primarily because it is a frozen world. Moreover, the terrestrial Arctic varies from area to area according to the location, glacial history, and climatic conditions. However, sites, networks, and opportunities for direct observations are still sparse relative to the warmer regions owing to physical and logistical limitations. To investigate the impact of climate change in this region, a number of studies using both analysis of observed data and numerical modelling have been carried out (e.g., Zhang et al., 2005; Brown and Robinson, 2011; Brutel-Vuilmet et al., 2013; Koven et al., 2011, 2013; Slater and Lawrence, 2013). Various numerical modelling schemes have been developed to treat physical and biogeochemical processes on and below the land surface. Some of these processes are site-specific or process-oriented, while others are implemented as components of atmosphere–ocean coupled global climate models (AOGCMs), or Earth system models (ESMs) to interact with the overlying atmosphere. Among these processes, snowpack, ground freezing/thawing, and carbon exchange are the most relevant and important processes in terrestrial process models (TPM) for investigating the climate and ecosystem of the pan-Arctic region.

1.1 GRENE-Arctic project and GTMIP

The GRENE-TEA model intercomparison project (GTMIP) was originally planned as part of the terrestrial research project of the GRENE Arctic Climate Change Research Project (GRENE-TEA) to achieve the following targets: a) to pass possible improvements regarding physical and biogeochemical processes for Arctic terrestrial modelling (excluding glaciers and ice sheets) in the existing AOGCM terrestrial schemes for the AOGCM research community, and b) to lay the foundations for the development of future-generation Arctic terrestrial models. The project, however, involves groups of researchers from different backgrounds/disciplines (e.g., physics/geophysics, glaciology, biogeochemistry, ecosystem, forestry) with a wide range of research methods (e.g., field observations, remote-sensing, numerical modelling), target domains (e.g., Northern Europe, Siberia, Alaska, Northern Canada) and scales (from site-level to Pan-Arctic). As is often the case, multi-disciplinary opportunities were limited, initially creating a considerable challenge for the project (Fig. 1a). Communications between groups (e.g., modelling and field studies, physical and ecosystem disciplines, process-oriented and large-scale modelling), if any, were inconclusive and sporadic. Observational practices and procedures (e.g., variables to measure, equipment to use, standard zero depth for ground measurements) were different among groups and disciplines, and lacked standardization. Although each individual group had the needs and intention to interact with other groups, the requisite collaboration could not be achieved. Opinions obtained in the early stages revealed hidden quests for possible collaborations for "observational data for driving and/or validating data", "use of numerical models to test empirical hypothesis gained at the field", "interpretation of observed phenomena", and "optimization of observation network strategies." As a result of this situation, the model intercomparison project was deliberately blueprinted to promote communication and understanding between modelling and empirical scientists, and among modellers: the GTMIP protocols and datasets are set to function as a hub for the groups involved in the project (Fig. 1b). It also aimed to enhance the standardization of observation practices among the GRENE-TEA observation sites, and to form a tight collaboration between the field and modelling communities, laying a cornerstone for creating the driving dataset (details of the Stage 1 driving data and their creation as a product of collaboration between modellers and field scientists are documented by Sueyoshi et al. [2015]).

1.2 Model intercomparison for the terrestrial Arctic

Since the 1990s, a number of model intercomparison projects (MIPs) have been carried out, focusing on the performance of TPMs, AOGCMs, and ESMs; examples include PILPS (Project for Intercomparison of Land-Surface Parameterization Schemes; Henderson-Sellers, 1993), SnowMIP (Snow Models Intercomparison Project; Etchevers et al. 2004; Essery et al. 2009), Potsdam NPP MIP (Potsdam Net Primary Production Model Intercomparison Project; Cramer et al., 1999), C4MIP (Coupled Climate–Carbon Cycle Model Intercomparison Project; Friedlingstein et al. 2006), CMIP5 (Coupled Model Intercomparison Project; Taylor et al. 2012), and MsTMIP (Multi-scale synthesis and Terrestrial Model Intercomparison Project; Huntzinger et al., 2013), to name a few.

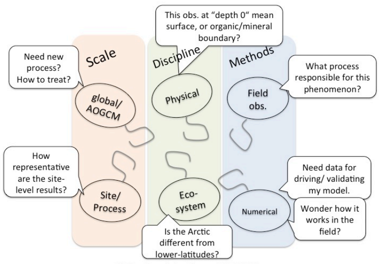
For snow dynamics, SnowMIP2 showed a broad variety in the maximum snow accumulation values, particularly at warmer sites and in warmer winters, although the duration of snow cover was relatively well simulated (Essery et al., 2009). The same study also noted that the SnowMIP2 models tend to predict winter soil temperatures that are too low in cold sites and for sites with shallow snow, a discrepancy arguably caused by the remaining uncertainties in ecological and physical processes and the scarcity of winter process measurements for model development and testing in the boreal zone. The CMIP5 models simulated the snow cover extent for most of the Arctic region well, except for the southern realm of the seasonal snow cover area (Brutel-Vulmet et al., 2013). The poor performance of some of the TPMs in this region is due to an incorrect timing of the snow onset, and possibly by an incorrect representation of the annual maximum snow cover fraction (Brutel-Vulmet et al., 2013). For ground freezing/thawing processes, Koven et al. (2013) showed the current status of the performance of AOGCMs for permafrost processes based on CMIP5 experiments. There was large disagreement among modelled soil temperatures, which may have been due to the representation of the thermal connection between the air and the land surface and, in particular, its mediation by snow in winter. Vertical profiles of the mean and amplitude of modelled soil temperatures showed large variations, some of which can be attributed to differences in the physical properties of the modelled soils and coupling between energy and water transfer. This appears to be particularly relevant for the representation of organic layers.

For the biogeochemical cycles, a number of studies based on MIPs have been carried out. The broad global distribution of net primary productivity (NPP) and the relationship of annual NPP to the major climatic variables coincide in most areas with differences among the 17 global terrestrial biogeochemical models that cannot be attributed to the fundamental modelling strategies (Cramer et al., 1999). The ESMs in CMIP5 use the climate and carbon cycle performance metrics, and they showed that the models correctly reproduced the main climatic variables controlling the spatial and temporal characteristics of the carbon cycle (Anav et al., 2013). However, several weaknesses were found in the modeling of the land carbon cycle: for example, the leaf area index is generally overestimated by models compared with remote sensing data (Anav et al., 2013); NPP and terrestrial carbon storage responses to CO2 increases greatly differs among models (Hajima et al., 2014); current ESMs displays large variations for the estimated soil carbon amounts, in particular for northern high latitudinal regions, and lack the capability to represent the potential degradation of frozen carbon in permafrost regions (Todd-Brown et al., 2014). The future projection by ESMs suggests that the carbon sink characteristic will increase in northern high latitudes, although there are some uncertainties, such as nutrient limitations in CO2 fertilization, the effect of soil moisture on decomposition rates, and mechanistic representations of permafrost (Qian et al., 2010; Ahlstrom et al., 2012; Arora et al., 2013). It should be noted that the reference observation data used for these evaluations are prone to uncertainties due to random and bias errors in the measurements themselves, sampling errors, and analysis error, especially for biogeochemical variables such as land gross primary productivity (GPP) (e.g., Anav et al., 2013; Piao et al., 2013). Based on the outcomes of these MIPs, TPMs have improved their performances.

At scales from a continental level (including those mentioned above) to a site level (model-observation comparisons; e.g., Zaehle et al., 2014), different MIPs have also been conducted, and generally study physical or ecosystem processes separately. PILPS (Henderson-Sellers et al., 1993) and a series of snow MIPs (Etchevers et al., 2004; Essery et al., 2009) are well-known MIPs for physical processes, targeting hydrology and snow dynamics. Recently, an MIP for tundra sites has been conducted, but its focus is limited to soil thermal dynamics (Ekici et al., 2014). In turn, ecosystem MIPs on continental scales have two predecessors: i.e., the North American Carbon Program Site Synthesis (Schwalm et al., 2010) and CarboEastAsia-MIP (Ichii et al., 2013). Although both MIPs employ multiple terrestrial biosphere models to different eddy-covariance measurement sites (Schwalm et al. (2010) with 22 models for 44 sites in North America; Ichii et al. (2013) with 8 models for 26 sites in East Asia), boreal and Arctic sites were not the major targets. In other studies targeting specific eco-climatic regions, the Arctic was again not the main domain: Jung et al. (2007) assessed GPPs for Europe, and Ichii et al. (2010) for Japan. Rawlins et al. (2015) assessed carbon budget differences among several GCM-compatible models in northern Eurasia, with little examination of the physical processes. In other regions than the Arctic, there have been cross-sectional evaluations of physical and ecosystem processes, such as Morales et al. (2005), evaluating carbon and water fluxes in Europe, and de Gonçalves et al. (2013), the LBA-Data Model Intercomparison Project (LBA-DMIP), analysing water and carbon fluxes in the Amazon.

The GTMIP consists of two stages (Fig. 2): one dimensional, historical GRENE-TEA site evaluations for examining the model’s behaviour and its uncertainty (Stage 1), and circumpolar evaluations using projected climate change data from GCM outputs (Stage 2). Hereafter, we describe the Stage 1 protocol. This stage aims to evaluate the physical and biogeochemical TPMs through three-decade site simulations driven and validated by the GRENE-TEA site-derived data. It calls for broader participation in the activity from a wider community to assure robust assessments for model-derived uncertainty, and to efficiently investigate the terrestrial system response to climate variability considering the diversity of the pan-Arctic sites. Thus, the scope and geographical domain of GTMIP Stage 1 is unique in its target of the Arctic region, including both taiga and tundra, and in its evaluations of the behaviour of the energy-snow-soil-vegetation subsystem, employing a wide range of models from physical land surface schemes to terrestrial ecosystems. ”

- New Figures 1 and 2:



“Pirates” before GTMIP

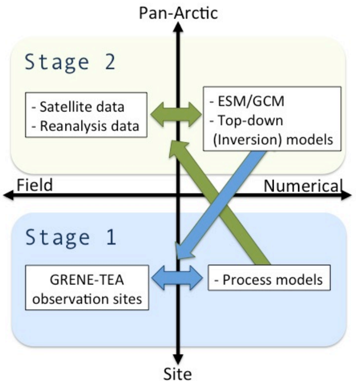


Figure 2. Schematic diagram for stage 1 and 2 of GTMIP

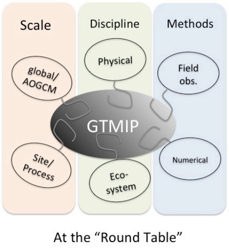


Figure 1. “Pirates of the Arctic” sit together at thea Round Table

- 2.6 Currently participating models

“Participation in GTMIP Stage 1 is voluntary and open to any interested modellers or institutions. 16 TPMs have announced their participation in GTMIP Stage 1. These models are...”

- 3. Analysis plan and exemplary results

“This section presents the analysis plan for GTMIP Stage 1 and sample outputs based on already submitted materials. To answer the key questions for the target processes proposed in Sect. 2.1, we plan to analyze the model output by describing the model–model and model–observation differences, discerning the cause of these differences, and investigating parameter sensitivity. The outputs of multiple models will be compared in terms of the metrics shown in Table 3. These metrics are divided into five categories (i.e., energy and water budget, snowpack, phenology, subsurface hydrological and thermal states, and carbon budget). For terrestrial climate simulations on the decadal scale, the most important outputs are the latent heat flux (energy and water budget) and the net ecosystem exchange (carbon budget). The latent heat flux (evapotranspiration) is the essential driver of precipitation inland at high latitudes owing to high rates of recycling (e.g., Dirmeyer et al., 2009; Saito et al. 2006). Net ecosystem exchange (NEE) plays a fundamental role in determining global CO_2 concentrations by determining whether a site forms a carbon source or sink (e.g. Abramowitz et al., 2008; Mcguire et al., 2012). NEE represents the net land–atmosphere CO_2 flux, and a positive NEE represents net loss of CO_2 from the land to the atmosphere (i.e., carbon source; Mcguire et al., 2012). Although NEE is commonly used for tower flux observations and some

TPMs, the net ecosystem production (NEP) is used in GTMIP for both the observed and simulated values because it is more widely used in non-biogeochemical communities. A positive (negative) value of NEP represents a carbon sink (source).

Analyses will be organized and conducted in the following manner. Topical analyses, constituting major subsets of the project outcomes, will evaluate characteristics of model performances and their inter-site variations within each of the above five categories, while cross-sectional analyses between categories will explore the functionality and strength of interactions between processes. These analyses will be utilized for mining crucial processes to improve the site-level TPMs as well as large-scale GCM/ESM components.

First, the focus will be on model output variability for both the inter-annual and the inter-decadal time scales, based on the output time series over more than 30 years. Inter-site differences will also be evaluated for the four GRENE-TEA sites in the Arctic region, each of which has distinct characteristics. The vegetation type for three of the four sites is forest (two evergreen conifer: FB and KV; one deciduous conifer: YK) and the remaining site is tundra (TK). Three sites (FB, TK, and YK) are in the permafrost region, while KV is underlain by seasonally frozen ground. Figures 5–8 show statistical summary comparisons of the model outputs by site (the land cover and soil type parameters used for the simulations are shown in Table 2), expressing inter-model variations for physical and biogeochemical models using box plots for four variables of the metrics mentioned above: the annual mean latent heat flux (Qle_total_an), the annual maximum snow depth (SnowDepth_max), the annual gross primary production (GPP_an), and the annual net ecosystem production (NEP_an), respectively. When observed values were available (i.e., latent heat flux for FB for 2011–2013 and YK for 1998, 2001, 2003, 2004, 2007, and 2008), they are shown by black dots.

Second, the cause or attributes of the differences among models, or between models and observations, will be explored by employing statistical evaluations such as multivariate analyses and time series analyses on the metrics and individual eco-climate variables. This will improve understanding of the interrelation between the incorporated processes in each model. Figure 9 shows an exemplary comparison of a seasonal transition in the snow-permafrost-vegetation sub-system, expressed similarly by box plots. The figure summarizes the average dates for (from bottom to top) the completion of snow melt, the thawing of the top soil layer, the start and end of greening, the freezing of the top soil layer, and the start of seasonal snow accumulation. A comparison of the timings of these events over years and sites will illustrate individual models' characteristic behaviour in seasonal transitions, and their strength regarding process interactions, in combination with ordinary multivariate analysis techniques.

Finally, sensitivity tests for the model parameters are planned to quantify the effect of parameter sensitivity on the models' reproducibility. "

- New Figure 9

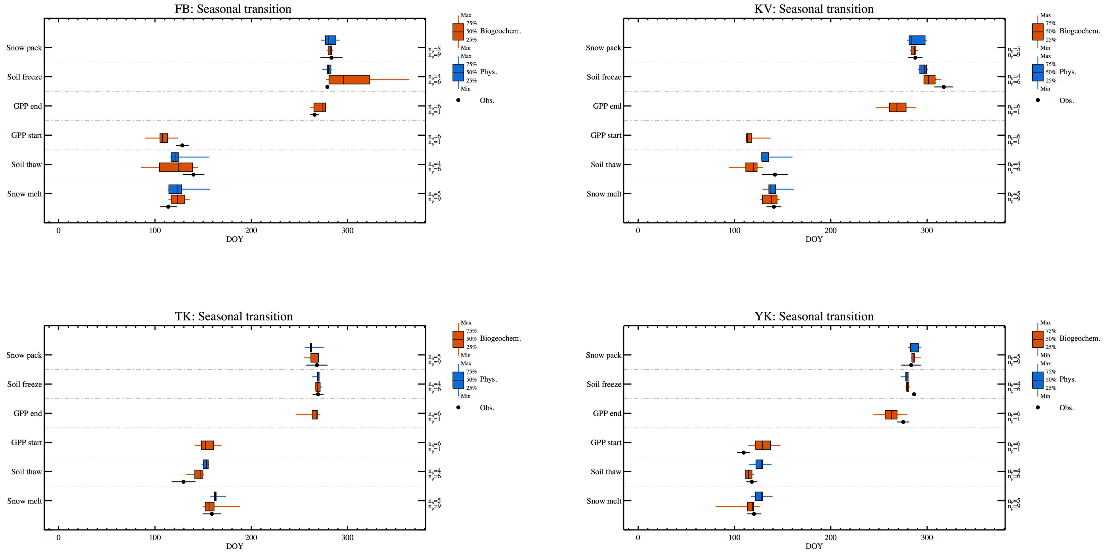


Figure 9. Example of seasonal transitions in ground temperature, snow, and vegetation among models.

<u>Techical comments:</u>

- *page (p.) 3, line (l.) 4: replace one of the "system" by another word*

We have revised to "the role and function of the terrestrial Arctic in the climate system"

- *p3., l5.: planned and conducted at the same ?*

We have removed "and being conducted" and reworded the part as "deliberatively designed."

- *p3, l9: replace "due to"by "using"*

We have replaced "due to" by "using."

- *p3. l20: delete "the"*

We have deleted “the.”

- *p4, l10: delete "some"*

We have deleted “some.”

- *p4, l20: replace "This" by "The same"*

We have replaced “This” by “The same.”

- *p4, l27: add a reference*

We have added the reference: Brutel-Vulmet et al., 2013

- *p6, l13: replace "This" by "a"*

We have replaced “This” by “A.”

- *p9, l6: "aims"*

We have changed to “aims.”

- *Caption Figure 3: "are shown by boxes"*

We have revised the text and a caption accordingly.

The GRENE-TEA Model Intercomparison Project (GTMIP): Overview and experiment protocol for Stage 1

S. Miyazaki[1, 2], K. Saito[2], J. Mori[1, 2], T. Yamazaki[3], T. Ise[4], H. Arakida[5], T. Hajima[2], Y. Iijima[1], H. Machiya[1, 2], T. Sueyoshi[1, 2], H. Yabuki[1, 2], E. J. Burke[6], M. Hosaka[7], K. Ichii[2], H. Ikawa[8], A. Ito[9], A. Kotani[12], Y. Matsuura[10], M. Niwano[7], T. Nitta[11], R. O'ishi[1, 11], T. Ohta[12], H. Park[2], T. Sasai[13], A. Sato[14], H. Sato[2], A. Sugimoto[15], R. Suzuki[2], K. Tanaka[2], S. Yamaguchi[14], K. Yoshimura[11]

[1]{National Institute of Polar Research, Japan}

[2]{Japan Agency for Marine-Earth Science and Technology, Japan}

[3]{Tohoku University, Japan}

[4]{Kyoto University, Japan}

[5]{RIKEN, Japan}

[6]{Met Office Hadley Centre, UK}

[7]{Meteorological Research Institute, Japan}

[8]{National Institute for Agro-Environmental Sciences, Japan}

[9]{National Institute for Environmental Studies, Japan}

[10]{Forestry and Forest products Research Institute, Japan}

[11]{The University of Tokyo, Japan}

[12]{Nagoya University, Japan}

[13]{University of Tsukuba, Japan}

[14]{National Research Institute for Earth Science and Disaster Prevention, Japan}

[15]{Hokkaido University, Japan}

Correspondence to:

K. Saito (ksaito@jamstec.go.jp)

Abstract

As part of the terrestrial branch of the Japan-funded Arctic Climate Change Research Project (GRENE-TEA), which aims to clarify the role and function of the ~~Arctic ~~terrestrial ~~system~~Arctic in the climate system~~,~~ and assess the influence of its changes on a global scale, this model intercomparison project (GTMIP) is ~~planned and being conducted~~deliberatively designed to 1) enhance communication and understanding between the "minds and hands" (i.e., between the modelling and field scientists) and 2) assess the uncertainty and variations stemming from variability in model implementation/~~designation~~design and in model outputs ~~due to~~using climatic and historical conditions in the Arctic terrestrial regions. This paper provides an overview of all GTMIP activity, and the experiment protocol of Stage 1~~ of the project,~~, which is site simulations driven by statistically fitted data created using the GRENE-TEA site observations for the last three decades. The target metrics for the model evaluation cover key processes in both physics and biogeochemistry, including energy budgets, snow, permafrost, phenology, and carbon budgets. ~~The preliminary~~Exemplary results ~~on~~for distributions of four metrics (annual mean latent heat flux, annual maximum snow depth, gross primary production, and net ecosystem production~~) already demonstrate the range of variations in reproducibility among extant models~~), and ~~sites. Full analysis on annual as well as ~~for seasonal ~~time scales,~~transitions are provided to ~~be conducted upon completion of model outputs submission,~~give an outlook of the planned analysis that will delineate the inter-dependence among the key processes, and provide ~~the clue~~clues for improving~~ the~~ model performance.

1 Introduction

The pan-Arctic ecosystem is characterized by low mean temperatures, snow cover, and seasonal frozen ground~~, and~~ or permafrost with a large carbon reservoir, covered by various biomes (plant types) ranging from deciduous and evergreen forests to tundra. The Arctic climate and ecosystem differ from the tropical and temperate counterparts primarily because it is a frozen world. Moreover, the terrestrial Arctic varies from area to area according to the location, glacial history, and climatic conditions. However, sites, networks, and opportunities for direct observations are still sparse relative to the warmer regions owing to physical and logistical limitations. To investigate the impact of climate change in this region, a number of studies using both analysis of observed data and numerical modelling have been carried out (e.g., Zhang et al., 2005; Brown and Robinson, 2011; Brutel-Vuilmet et al., 2013; Koven et al., 2011, 2013; Slater and Lawrence, 2013). Various numerical modelling schemes have been developed to treat physical and biogeochemical processes on and below the land surface~~, and interactions with the overlying atmosphere~~. Some of these processes are site-specific or process-oriented, while others are implemented as components of atmosphere–ocean coupled global climate models (AOGCMs), or Earth system models (ESMs~~).~~) to interact with the overlying atmosphere. Among these processes, snowpack, ground freezing/thawing, and carbon exchange are the most relevant and important processes in terrestrial process models (TPM) ~~applied in~~for investigating the climate and ecosystem of the pan-Arctic region.

1.1 GRENE-Arctic project and GTMIP

The GRENE-TEA model intercomparison project (GTMIP) was originally planned as part of the terrestrial research project of the GRENE Arctic Climate Change Research Project (GRENE-TEA) to achieve the following targets: a) to pass possible improvements regarding physical and biogeochemical processes for Arctic terrestrial modelling (excluding glaciers and ice sheets) in the existing AOGCM terrestrial schemes for the AOGCM research community, and b) to lay the foundations for the development of future-generation Arctic terrestrial models. The project, however, involves groups of researchers from different backgrounds/disciplines (e.g., physics/geophysics, glaciology, biogeochemistry, ecosystem, forestry) with a wide range of research methods (e.g., field observations, remote-sensing, numerical modelling), target domains (e.g., Northern Europe, Siberia, Alaska, Northern Canada) and scales (from site-level to Pan-Arctic). As is often the case, multi-disciplinary opportunities were limited, initially creating a considerable challenge for the project (Fig. 1a). Communications between groups (e.g., modelling and field studies, physical and ecosystem disciplines, process-oriented and large-scale modelling), if any, were inconclusive and sporadic. Observational practices and procedures (e.g., variables to measure, equipment to use, standard zero depth for ground measurements) were different among groups and disciplines, and lacked standardization. Although each individual group had the needs and intention to interact with other groups, the requisite collaboration could not be achieved. Opinions obtained in the early stages revealed hidden quests for possible collaborations for “observational data for driving and/or validating data”, “use of numerical models to test empirical hypothesis gained at the field”, “interpretation of observed phenomena”, and "optimization of observation network strategies." As a result of this situation, the model intercomparison project was deliberately blueprinted to promote communication and understanding between modelling and empirical scientists, and among modellers: the GTMIP protocols and datasets are set to function as a hub for the groups involved in the project (Fig. 1b). It also aimed to enhance the standardization of observation practices among the GRENE-TEA observation sites, and to form a tight collaboration between the field and modelling communities, laying a cornerstone for creating the driving dataset (details of the Stage 1 driving data and their creation as a product of collaboration between modellers and field scientists are documented by Sueyoshi et al. [2015]).

1.2 Model intercomparison for the terrestrial Arctic

Since the 1990s, a number of model intercomparison projects (MIPs) have been carried out, focusing on the performance of TPMs, AOGCMs, and ESMs; ~~some ~~examples include PILPS (Project for Intercomparison of Land-Surface Parameterization Schemes; Henderson-Sellers, 1993), SnowMIP (Snow Models Intercomparison Project; Etchevers et al. 2004; Essery et al. 2009), Potsdam NPP MIP (Potsdam Net Primary Production Model Intercomparison Project; Cramer et al., 1999), C4MIP (Coupled Climate–Carbon Cycle Model Intercomparison Project; Friedlingstein et al. 2006), CMIP5 (Coupled Model Intercomparison Project; Taylor et al. 2012), and MsTMIP (Multi-scale synthesis and Terrestrial Model Intercomparison Project; Huntzinger et al., 2013~~).~~), to name a few.

For snow dynamics, SnowMIP2 showed a broad variety in the maximum snow accumulation values, particularly at warmer sites and in warmer winters, although the duration of snow cover was relatively well simulated (Essery et al., 2009). ~~This~~The same study also noted that the SnowMIP2 models tend to predict winter soil temperatures that are too low in ~~the ~~cold sites and for sites with shallow snow, a discrepancy arguably caused by the remaining uncertainties in ecological and physical processes and the scarcity of winter process measurements for model development and testing in the boreal zone. The CMIP5 models simulated the snow cover extent for most of the Arctic region well, except for the southern realm of the seasonal snow cover area~~.~~ (Brutel-Vulmet et al., 2013). The poor performance of some of the TPMs in this region is due to an incorrect timing of the snow onset, and possibly by an incorrect representation of the annual maximum snow cover fraction (Brutel-Vulmet et al., 2013). For ground freezing/thawing processes, Koven et al. (2013) showed the current status of the performance of AOGCMs for permafrost processes based on CMIP5 experiments.

There was large disagreement among modelled soil temperatures, which may have been due to the representation of the thermal connection between the air and the land surface and, in particular, its mediation by snow in winter. Vertical profiles of the mean and amplitude of modelled soil temperatures showed large variations, some of which can be attributed to differences in the physical properties of the modelled soils and coupling between energy and water transfer. This appears to be particularly relevant for the representation of organic layers.

For the biogeochemical cycles, a number of studies based on MIPs have been carried out. The broad global distribution of net primary productivity (NPP) and the relationship of annual NPP to the major climatic variables coincide in most areas with differences among the 17 global terrestrial biogeochemical models that cannot be attributed to the fundamental modelling strategies (Cramer et al., 1999). The ESMs in CMIP5 use the climate and carbon cycle performance metrics, and they showed that the models correctly reproduced the main climatic variables controlling the spatial and temporal characteristics of the carbon cycle (Anav et al., 2013). ~~However, they found a weakness in the modeling of the land carbon cycle: a general overestimation of photosynthesis and leaf area index due to the lack of nutrient limitation on gross primary production (GPP).~~However, several weaknesses were found in the modeling of the land carbon cycle: for example, the leaf area index is generally overestimated by models compared with remote sensing data (Anav et al., 2013); NPP and terrestrial carbon storage responses to CO_2 increases greatly differs among models (Hajima et al., 2014); current ESMs displays large variations for the estimated soil carbon amounts, in particular for northern high latitudinal regions, and lack the capability to represent the potential degradation of frozen carbon in permafrost regions (Todd-Brown et al., 2014). The future projection by ESMs suggests that the carbon sink characteristic will increase in northern high latitudes, although there are some uncertainties, such as nutrient limitations in CO_2 fertilization, the effect of soil moisture on decomposition rates, and mechanistic representations of permafrost (Qian et al., 2010; Ahlstrom et al., 2012~~). As for the carbon-concentration feedback, the carbon cycle response to atmospheric CO_2 decreases for both the land and the ocean as CO_2 increases, which is related to saturation of the CO_2 fertilization effect and increased ecosystem respiration fluxes as vegetation and soil carbon biomass increase (~~; Arora et al., 2013). It should be noted that the reference observation data~~, which were~~ used for ~~those~~these evaluations~~,~~ are prone to uncertainties due to random and bias errors in the measurements themselves, sampling errors, and analysis error, especially for ~~the~~ biogeochemical variables such as land gross primary productivity (GPP) (e.g., Anav et al., 2013; Piao et al., 2013). Based on the outcomes of these MIPs, TPMs have improved their performances.

At scales~~Based on the outcomes of these MIPs, TPMs have improved their performances. However, as past MIPs were carried out on a global scale or in the subarctic region using gridded outputs~~ from ~~the models, intercomparisons dedicated~~a continental level (including those mentioned above) to ~~Arctic region processes that include both physical and biogeochemical aspects at~~ a site level ~~are still limited (e.g. Ekici et al., 2014; Rawlins et al., 2015; Wang et al., 2015). A mission of the modelling group in the terrestrial research project of the GRENE Arctic Climate Change Research Project (GRENE-TEA) is to a) pass possible improvements regarding physical and biogeochemical processes for Arctic terrestrial modelling (excluding glaciers and ice sheets) in the extant AOGCM terrestrial schemes to the AOGCM research community, and b) lay the foundations for the development of future-generation Arctic terrestrial models. This~~ (model~~ intercomparison project (GTMIP) is planned and being conducted to achieve these goals. It is also designed to promote communication and understanding between modelling and empirical scientists, to assess the effect of model implementation on model uncertainty and variations, and to investigate the model output variability due to climatic and historical conditions among the pan-Arctic sites.~~ -observation comparisons; e.g., Zaehle et al., 2014), different MIPs have also been conducted, and generally study physical or ecosystem processes separately. PILPS (Henderson-Sellers et al., 1993) and a series of snow MIPs (Etchevers et al., 2004; Essery et al., 2009) are well-known MIPs for physical processes, targeting hydrology and snow dynamics. Recently, an MIP for tundra sites has been conducted, but its focus is limited to soil thermal dynamics (Ekici et al., 2014). In turn, ecosystem MIPs on continental scales have two predecessors: i.e., the North American Carbon Program Site Synthesis (Schwalm et al., 2010) and CarboEastAsia-MIP (Ichii et al., 2013). Although both MIPs employ multiple terrestrial biosphere models to different eddy-covariance measurement sites (Schwalm et al. (2010) with 22 models for 44 sites in North America; Ichii et al. (2013) with 8 models for 26 sites in East Asia), boreal and Arctic sites were not the major targets. In other studies targeting specific eco-climatic regions, the Arctic was again not the main domain: Jung et al. (2007) assessed GPPs for Europe, and Ichii et al. (2010) for Japan. Rawlins et al. (2015) assessed carbon budget differences among several GCM-compatible models in northern Eurasia, with little examination of the physical processes. In other regions than the Arctic, there have been cross-sectional evaluations of physical and ecosystem processes, such as Morales et al. (2005), evaluating carbon and water fluxes in Europe, and de Gonçalves et al. (2013), the LBA-Data Model Intercomparison Project (LBA-DMIP), analysing water and carbon fluxes in the Amazon.

The GTMIP consists of two stages~~:~~ (Fig. 2): one dimensional, historical GRENE-TEA site evaluations for examining the model's behaviour and its uncertainty (Stage 1~~)~~), and circumpolar evaluations using projected climate change data from GCM outputs (Stage 2). ~~This paper focuses on~~ Hereafter, we describe the Stage 1 ~~of the project, which evaluates the TPMs for~~protocol. This stage aims to evaluate the physical and biogeochemical ~~processes by~~TPMs through three-decade site simulations ~~for the last three decades,~~ driven and validated by the GRENE-TEA site~~ observation~~-derived data~~ that were compiled through~~ . It calls for broader participation in the activity from a ~~tight collaboration between the GRENE-TEA field and modelling groups.~~wider community to assure robust assessments for model-derived uncertainty, and to efficiently investigate the terrestrial system response to climate variability considering the diversity of the pan-Arctic sites. Thus, the scope and geographical domain of GTMIP Stage 1 is unique in its target of the Arctic region, including both taiga and tundra, and in its evaluations of the behaviour of the energy-snow-soil-vegetation subsystem, employing a wide range of models from physical land surface schemes to terrestrial ecosystems.

2 Experiment design

2.1 Targeted processes

In GTMIP, a variety of models ranging from specific models that focus on snowpack formation processes to highly complex DGVMs are expected to participate. The following five categories (from "a" to "e" below) ~~were selected as~~set the unit for the key processes to assess the performance of the ~~extant~~existing TPMs in the pan-Arctic region, to evaluate the variations among the models and the mechanisms behind their strengths and weaknesses, and to obtain information and guidance to improve the next generation of TPMs. The five categories are a) exchange of energy and water between atmosphere and land, b) the snowpack, c) phenology, d) ground freezing/thawing and the active layer, and e) the carbon budget. The categories cover the essential processes that make the pan-Arctic region unique compared with other regions: seasonal changes in both physical and biogeochemical processes and the associated strong climate feedback, which are characterized by liquid-ice phase changes, the subsequent ecosystem response, and their interactions.

The scientific questions at the Stage 1 are: How well do the TPMs reproduce target metrics (examples are shown in column B in Table 1) in terms of agreement with observations? How do the reproductions vary among the models? If the reproductions are good or poor in some models, which processes in the TPMs are responsible and why?

2.2 Driving datasets and model parameters

The target period for Stage 1 was set from 1980 to 2013~~, providing~~ to provide at least 30 years of data~~ to enable~~, the minimum requirement for climatological analyses. The period is also favourable in terms of the accuracy and coherence of the relevant large-scale climate data thanks to the fully fledged operation of various satellite observations (e.g., Dee et al., 2011). We ~~provided~~are providing the following driving data for Stage 1: surface air temperature, precipitation, specific humidity, air pressure, wind speed, incident short-wave and long-wave radiation.

For this stage (site simulations), forcing and validation data have been prepared, taking maximum advantage of the observation data from GRENE-TEA sites in operation (Fairbanks (FB) in Alaska; Tiksi (TK), Yakutsk (YK), Chokurdakh (CH), and Tura (TR) in Russia; and Kevo (KV) in Finland, shown in Fig. ~~1), to evaluate the inter-model and inter-site variations for 1980–2013. The backbone of the continuous forcing data (called “level 0” or L0; Saito et al., 2014a) was constructed from reanalysis data to avoid limited coverage and/or missing data, or the lack of consistency inherent in observational data, with bias-corrected monthly Climate Research Unit (CRU) for temperature; Harris et al., 2014 and Global Precipitation Climatology Project (GPCP) for precipitation; Adler et al., 2003 datasets at the respective nearest grid to the sites. The European centre for medium-range weather forecasts ReAnalysis (ERA)-interim reanalysis data (Dee et al., 2011) were chosen from four products (National Centers for Environmental Prediction (NCEP)/ National Center for Atmospheric Research (NCAR); NCEP/NCAR, NCEP-Department of Energy (DOE), Japanese Reanalysis (JRA)-55, and ERA-interim) because it showed the smallest bias relative to the monthly CRU and GPCP in terms of 2-m air temperature and precipitation in the pan-Arctic region (north of 60°N).~~

3), to evaluate the inter-model and inter-site variations for 1980–2013. These sites, the latitude of which varies from 62°N–71°N, have different characteristics in terms of climate (e.g., air temperature, precipitation), snow (e.g., type, amount and accumulation period), vegetation, and frozen ground conditions (Sueyoshi et al., 2015), providing a good representation of the diversity of the terrestrial Arctic.~~Assimilation of the observed data was then applied to reflect local characteristics and to derive the primary driving data, "level 1" data (L1; Saito et al., 2014b) and, in addition, the level 1 hybrid data (L1H) by replacing data with observed data when available.~~ ~~The L1 dataset was provided for four sites (FB, KV, TK and YK) due to availability of observed data for validations. Further details of the method used to create the L0 and L1 datasets, and their basic statistics, are described in Sueyoshi et al. (2015, to be submitted).~~

~~The 20-year detrended meteorological driving dataset was provided for spin up, allowing biogeochemical models to set up initial soil carbon conditions without including warming trends and/or ENSO (El Niño Southern Oscillation).~~ ~~This dataset is based on the L1 data for the period of 1980–1999 (Saito et al., 2015). The monthly values of the photosynthetically active radiation (fPAR) and leaf area index (LAI) datasets at GRENE-TEA sites, created based on Moderate Resolution Imaging Spectroradiometer (MODIS) satellite data (MOD15A2, MYD15A2), were also provided where required (Saito et al., 2014c).~~

~~The driving datasets are provided in the ASCII fixed-length record files, and are available through the Arctic Data Archive System (ADS; https://ads.nipr.ac.jp/gtmip/gtmip.html), along with the simulation protocol.~~

~~The site description, including location, dominant vegetation type, soil, climate, fPAR, LAI, data available for model validation, and references for observation data, is summarized in Table 2.~~ The annual air temperature and precipitation at the six sites ranges from –13.5 °C to –1.6 °C and from 188 mm to 415 mm, respectively. Four sites (FB, KV, YK, and TR) are in the boreal forest, while TK is in tundra and CH in the tundra–forest transition zone. Most of the sites are located in the permafrost zone with an active layer ranging from 0.4 m to 1.2 m, except for the KV site, which is seasonally frozen.

Because of the severe conditions for maintaining monitoring sites in arctic region, continuous observation data over years are scarce, which makes it very difficult to create ready-to-drive data directly from observations (e.g., owing to missing values, discontinuity of measurement periods, outliers). To overcome this problem, we first constructed the backbone of the continuous forcing data (called “level 0” or L0; Saito et al., 2014a) from climate reanalysis products to avoid the issues of limited coverage and/or missing data, or the lack of consistency inherent in observational data, using the bias-corrected monthly Climate Research Unit (CRU) for the temperature dataset (Harris et al., 2014) and the Global Precipitation Climatology Project (GPCP) for the precipitation dataset (Adler et al., 2003) at the respective nearest grid to the sites. The European Centre for Medium-range Weather Forecasts ReAnalysis (ERA)-interim reanalysis data (Dee et al., 2011) were chosen from four products (National Centers for Environmental Prediction (NCEP)/ National Center for Atmospheric Research (NCAR); NCEP/NCAR, NCEP-Department of Energy (DOE), Japanese Reanalysis (JRA)-55, and ERA-interim) because they showed the smallest bias relative to the monthly CRU and GPCP in terms of 2-m air temperature and precipitation in the pan-Arctic region (north of 60°N).

Assimilation of the observed data was then applied to reflect local characteristics and to derive the primary driving data, “level 1” data (L1; Saito et al., 2014b) and, in addition, the level 1 hybrid data (L1H) by replacing data with observed data when available. The L1 dataset was provided for four sites (FB, KV, TK, and YK) owing to the availability of the observed data for validations. For the creation of the site-specific data, collaboration with the field scientists who are in charge of the observation sites and know the circumstances of the data obtained was critical. Further details on the creation of the L0 and L1 datasets, and their basic statistics, are described in Sueyoshi et al. (2015).

As the warming trend is becoming visible, in particular for northern high-latitude regions (IPCC, 2013), the 20-year detrended meteorological driving dataset is provided for spin up, allowing biogeochemical models to set up initial soil carbon conditions without the warming trends and/or ENSO (El Niño Southern Oscillation). This dataset is based on the L1 data for the period of 1980–1999 (Saito et al., 2015). The monthly values of the photosynthetically active radiation (fPAR) and leaf area index (LAI) datasets at GRENE-TEA sites, created based on Moderate Resolution Imaging Spectroradiometer (MODIS) satellite data (MOD15A2, MYD15A2), are also provided where required (Saito et al., 2014c). These driving datasets are provided in the ASCII fixed-length record files, and are available through the Arctic Data Archive System (ADS; https://ads.nipr.ac.jp/gtmip/gtmip.html), along with the simulation protocol.

The site description, including locations, dominant vegetation types, soil, climate, fPAR, LAI, data for model validation, and references for observation data, is summarized in Table 2.

2.3 Model setup

As already proposed in existing MIP studies (e.g., Ichii et al., 2010), we set Stage 1 ~~consists~~to consist of two further sub-stages: 1A and 1B. Stage 1A, which ~~aim~~aims to evaluate the inter-model variations in baseline performance at each site, requested the participants to use the parameters in the default settings for the provided boundary conditions, such as land cover type. In contrast, Stage 1B allows tuning for the best reproduction of observations so that the parameter sensitivity among the sites can be evaluated. Process 1B is particularly important for the pan-Arctic region because many monitoring sites are located in temperate regions and models are generally validated against these environmental conditions.

We set the initial condition date to 01 September 1979, so that simulations started with a no-snow condition. The initial data for the model boundary conditions ~~were~~are available, as most stations can provide observation data for soil temperature and soil moisture profiles. However, each model could use its own method for initialization.

The spin up process may also differ between models. However, we ~~recommended~~recommend continuing spin up until a steady state ~~was~~is achieved for the main variables (see Sect. 2.5). For example, Takata (2002) defined ~~the~~a threshold of a steady state in a slowly varying system as

~~$\frac{X_n - X_{n-1}}{X_n} < 10^{-2}$~~

$$\frac{X_n - X_{n-1}}{X_n} < 10^{-2}$$

(1)

where X is a physical variable (e.g., fluxes, ground temperature, soil moisture, or ice content). The subscript n denotes the annual mean for the n-th year.

For biogeochemical cycle models, ~~we recommended~~in particular, we recommend maintaining spin up over at least 2000 years using the detrended meteorological driving data (also provided through ADS) because soil accumulation is quite slow owing to the low soil temperature, and pre-industrial atmospheric CO_2 concentrations (e.g., 280 ppmv ~~for~~ around the year ~~of~~ 1750~~–~~) until the soil carbon reached equilibrium; the atmospheric CO_2

concentration should then be increased to the current level (e.g., 340 ppmv) over 200 years or so (the period being dependent on the model). For the submission period (1979 to 2013), use of the historical atmospheric CO_2 concentration ~~was~~is recommended for these models so that they are driven by time-variant CO_2 concentrations.

~~2.4 Analysis plan~~

~~To answer the key questions for the target processes proposed in Sect. 2.1, we planned to analyze the model output by describing the model–model and model–observation differences, discerning the cause of differences, and investigating parameter sensitivity. We compared the outputs of multiple models using the metrics shown in Table 3. These metrics, divided into five categories (i.e., energy and water budget, snowpack, phenology, subsurface hydrological and thermal states, carbon budget), were used to evaluate model performance in each category, and search for clues to improve the TPMs. For the decadal-scale climate simulation, the most important outputs are the latent heat flux and the net ecosystem exchange. The latent heat flux (evapotranspiration) is the essential driver of precipitation inland at high latitudes owing to high rates of recycling (e.g., Dirmeyer et al., 2009; Saito et al. 2006). Net ecosystem exchange (NEE) plays a fundamental role in determining global CO_2 concentrations by determining whether a site forms a carbon source or sink (e.g. Abramowitz et al., 2008; Meguire et al., 2012). NEE represents the net land–atmosphere CO_2 flux, and a positive NEE represents net loss of CO_2 from the land to the atmosphere (i.e., carbon source; Meguire et al., 2012). Although NEE is commonly used for the tower flux observation and some TPMs, we decided to use the net ecosystem production (NEP) for both the observed and simulated values because the latter is more widely used in non-biogepchemical communities. A positive (negative) value of NEP represents a carbon sink (source).~~

~~First, we will examine both the inter-annual and the inter-decadal model output variability based on the output time series over more than 30 years. Inter-site differences will also be evaluated for the four GRENE-TEA sites in the Arctic region, each of which has distinct characteristics. The vegetation type for three of the four sites is forest (two evergreen conifer: FB and KV; one deciduous conifer: YK) and the remaining site is tundra (TK). Three sites (FB, TK, and YK) are in the permafrost region, while KV is underlain by seasonally frozen ground.~~

~~Second, we aim to discern the cause of differences among models, or between models and observations, by employing statistical evaluations such as multivariate analyses and time series analyses to investigate the connections between the metrics and individual eco-climate variables. This will improve understanding of the interrelation between the incorporated processes in each model. Finally, we will conduct sensitivity tests for the model parameters to quantify the effect of parameter sensitivity on models' reproducibility.~~

~~2.5~~2.4 Model output variables

We ~~are requesting~~request participants to submit those variables listed in Table S1 (refer to the Supplementary Material) in ASCII format with CSV-type files. The template file for output submission has been provided through ADS. ~~The file naming convention for submitting the result of each model is defined as follows.~~

~~[Model-ID]_[stage-ID]_[forcing ID]_[station-ID]_[yymmdd (date of submission)].csv,~~

~~where stage_ID is either "1a" or "1b," forcing_ID is "L0," "L1," or "L1H," and station_ID is shown in Table S2 (refer to the Supplementary Material).~~

The variables for submission are categorized into six groups: 0) model driving, 1) energy and water budget, 2) snow dynamics, 3) vegetation, 4) subsurface hydrological and thermal states, and 5) carbon budget~~. The~~, in parallel to the analysis categories. Since the spectrum of the participating models is expected to be very large (ranging from physical to biogeochemical to ecosystem models; Fig. 4), we made an extensive list of output variables to cover the expected range. However, the actual output variables a model submits will be dependent on the model's specification. Considering this spread, the priority for each variable, classed at three levels, was set according to the necessity and availability for evaluation of the model performance. In addition, participants are requested to provide information on the status of the variables in their model (i.e., model driving, prescribed parameter, prognostic, diagnostic, or not applicable), through the provided questionnaire (Supplementary Material, Table S3; provided through ADS), to identify the characteristics of the model.

Although the temporal resolution of a variable should depend on the model, we ~~are requesting~~request submission of the variables with the minimum temporal resolution available for the model. For the models that provide daily outputs, the time for each day should be defined by the local time (FB: UTC – 10; KV: UTC + 2; TK: UTC + 9; YK: UTC + 9; CH: UTC + 10; TR: UTC + 7). Those models that use the no-leap calendar (365 days for all years) are requested to leave out 29 February. For those models with a 360-day calendar, data on Days of Year (DOYs) 90, 151, 212, 304, and 365 (corresponding to March 31, May 31, July 31, October 31, and December 31 in a no-leap year) should be omitted.

~~2.6 Participating models~~

2.5 ~~As of 01 March 2015, 16 TPMs have been~~Currently participating models

Participation in ~~the~~ GTMIP~~,~~ Stage 1~~. Those~~ is voluntary and open to any interested modellers or institutions. 16 TPMs have announced their participation in GTMIP Stage 1. These models are the permafrost model (FROST), physical snow models (SMAP and SNOWPACK), land surface models (2LM, HAL, JULES, several versions of MATSIRO, and SPAC-Multilayer), a physical and biogeochemical soil dynamics model (PB-SDM), terrestrial biogeochemical models (BEAMS, Biome-BGC, STEM1, and VISIT), dynamic global vegetation models (LPJ and SEIB-DGVM, coupled with a land surface model [Noah-LSM] or stand-alone), and a coupled hydrological and biogeochemical model (CHANGE). The models with higher degrees of complexity in their treatment of physical processes are 2LM, CHANGE, FROST, HAL, JULES, MATSIRO, PB-SDM, SNOWPACK, SMAP, and SPAC-multilayer. The models with higher degrees of complexity in their treatment of biogeochemical processes are BEAMS, Biome-BGC, CHANGE, LPJ, SEIB-DGVM, STEM1, and VISIT. The models enabled to couple with AOGCMs (currently, JULES, HAL, LPJ, MATSIRO, and SMAP) make up about 30% of the participating models.

To illustrate the variability of the participating models with respect to the implemented physical and biogeochemical processes, we created a diagram showing the habitat of the currently participating models (Fig. ~~2~~4) by incorporating the model survey results referred to in the previous section. The spread of the ~~currently participating~~ models is large for both physical ~~processes~~ and biogeochemical ~~processes~~process dimensions, which will benefit the evaluation and attribute examinations of the models regarding their ability to reproduce observations.

3 ~~Preliminary~~Analysis plan and exemplary results

This section presents ~~preliminary results~~ the analysis plan for GTMIP Stage 1 and sample outputs based on ~~the outputs~~ already submitted ~~for~~ materials. To answer the key questions for the target processes proposed in Sect. 2.1, we plan to analyze the model output by describing the model–model and model–observation differences, discerning the cause of these differences, and investigating parameter sensitivity. The outputs of multiple models will be compared in terms of the metrics shown in Table 3. These metrics are divided into five categories (i.e., energy and water budget, snowpack, phenology, subsurface hydrological and thermal states, and carbon budget). For terrestrial climate simulations on the decadal scale, the most important outputs are the latent heat flux (energy and water budget) and the net ecosystem exchange (carbon budget). The latent heat flux (evapotranspiration) is the essential driver of precipitation inland at high latitudes owing to high rates of recycling (e.g., Dirmeyer et al., 2009; Saito et al. 2006). Net ecosystem exchange (NEE) plays a fundamental role in determining global CO_2 concentrations by determining whether a site forms a carbon source or sink (e.g. Abramowitz et al., 2008; Mcguire et al., 2012). NEE represents the net land–atmosphere CO_2 flux, and a positive NEE represents net loss of CO_2 from the land to the atmosphere (i.e., carbon source; Mcguire et al., 2012). Although NEE is commonly used for tower flux observations and some TPMs, the net ecosystem production (NEP) is used in GTMIP for both the observed and simulated values because it is more widely used in non-biogeochemical communities. A positive (negative) value of NEP represents a carbon sink (source).

Analyses will be organized and conducted in the following manner. Topical analyses, constituting major subsets of the project outcomes, will evaluate characteristics of model performances and their inter-site variations within each of the above five categories, while cross-sectional analyses between categories will explore the functionality and strength of interactions between processes. These analyses will be utilized for mining crucial processes to improve the site-level TPMs as well as large-scale GCM/ESM components.

First, the focus will be on model output variability for both the inter-annual and the inter-decadal time scales, based on the output time series over more than 30 years. Inter-site differences will also be evaluated for the four GRENE-TEA sites in the Arctic region, each of which has distinct characteristics. The vegetation type for three of the four sites is forest (two evergreen conifer: FB and KV; one deciduous conifer: YK) and the remaining site is tundra

(TK). Three sites (FB, TK, and YK) are in the permafrost region, while KV is underlain by seasonally frozen ground. ~~Stage 1A, in which~~ Figures 5–8 show statistical summary comparisons of the model outputs by site (the land cover and soil type parameters ~~are kept at~~used for the ~~default settings~~ simulations are shown in Table 2~~. In this paper, we have focused on the~~), expressing inter-model variations for physical and biogeochemical models using box plots for four variables of the metrics mentioned ~~in Sect. 2.4:~~above: the annual mean latent heat flux (Qle_total_an), the annual maximum snow depth (SnowDepth_max), the annual gross primary production (GPP_an), and the annual net ecosystem production (NEP_an~~).~~), respectively. When observed values were available (i.e., latent heat flux for FB for 2011–2013 and YK for 1998, 2001, 2003, 2004, 2007, and 2008), they are shown by black dots.

Second, the cause or attributes of the differences among models, or between models and observations, will be explored by employing statistical evaluations such as multivariate analyses and time series analyses on the metrics and individual eco-climate variables. This will improve understanding of the interrelation between the incorporated processes in each model.

~~3.1 Latent heat flux and annual maximum snow depth~~

~~Annual mean latent heat flux is one of the best metrics for evaluating the energy and water budget reproducibility of TPMs for annual time scales. Figure 3 shows a comparison of the model outputs by site, expressing intra-model variations by box plots. When observed values were available (i.e., for FB for 2011–2013 and YK for 1998, 2001, 2003, 2004, 2007, and 2008), they are shown by black dots. The physical-processes-oriented models (hereafter, P-models: 2LM, JULES, MATSIRO, and PB-SDM) generally reproduced observed latent heat flux well at FB and YK, while the biogeochemical-processes-oriented models (hereafter, BG-models: BEAMS (only for 2001–2011), Biome-BGC, CHANGE, SEIB-DGVM, and VISIT) tended to show higher values than the observations. The inter-model variation of Qle_total_an among BG-models was higher than among P-models at KV and TK, where it was not possible to compare the model output with data since no flux observations were conducted.~~

~~Annual maximum snow depth is an important metric for evaluating the snowpack process, especially the snow accumulation process, and the water resources in TPMs. Figure 4 shows a similar comparison to Fig. 3 for maximum snow depth. Note that for those models that~~

~~calculate snow water equivalent (SWE) but not snow depth (SD), we converted SWE to SD assuming a constant snow density of 300 kg m^{-3}. Snow observations were conducted at all sites: for 1980–2012 at FB, 1996–2013 at KV, 1980–2008 at TK, and 1980–2008 at YK. While dedicated snow models are good at reproducing maximum snow depth (except for at TK), those models with an invariant snow density show relatively lower values than the observations. Both P-models (2LM, JULES, MATSIRO, PB-SDM, and SMAP for snow) and BG-models gave values that were approximately double the observed values of the SnowDepth_max at TK. This overestimation of SnowDepth_max at TK is probably related to the formation of wind crust or snowdrift by strong winds, as noted by Hirashima et al., (2004).~~

~~3.2 Carbon budget~~

~~Annual gross primary production is a good indicator for evaluating the photosynthesis process in TPMs. Figure 5 shows a similar analysis to Fig. 3, but for GPP_an from BG-models (BEAMS, Biome-BGC, CHANGE, LPJ, SEIB-DGVM, STEM1, and VISIT for carbon budget). The observed values for the carbon budget are available only at FB (2011–2013) and YK (2004–2012). The simulated GPP_an for FB covers a wide range of values, from half the observed value of 0.4 kg C m^{-2} yr^{-1} to more than three times the observed value. The simulated GPP_an for YK showed smaller variations both among models and between models and observed values, from half the observed value of 0.6 kg C m^{-2} yr^{-1} up to 0.7 kg C m^{-2} yr^{-1}. Results for KV, showed similar variations (50%) to FB, with no outliers. Results for TK show the smallest absolute values, with small inter-model variations, which are probably due to the characteristics of the tundra vegetation.~~

~~Annual net ecosystem production is a substantial parameter in determining whether a site is a carbon source or sink. Figure 6 shows a similar analysis to Fig. 5, but for NEP_an. The simulated NEP_an was positive at all sites except one outlier at FB, which suggests all the examined sites are carbon sinks. The simulated NEP_an for FB was about 0.03–0.06 kg C m^{-2} yr^{-1}, while the observed value was almost zero. Note that the value for the observation at FB is only derived from 2011–2012, while that for simulation is a 34-year average. The simulated NEP_an for YK was about half of the observed value. Generally the inter-model range of NEP_an was smaller than that of GPP_an.~~

Figure 9 shows an exemplary comparison of a seasonal transition in the snow-permafrost-vegetation sub-system, expressed similarly by box plots. The figure summarizes the average dates for (from bottom to top) the completion of snow melt, the thawing of the top soil layer, the start and end of greening, the freezing of the top soil layer, and the start of seasonal snow accumulation. A comparison of the timings of these events over years and sites will illustrate individual models' characteristic behaviour in seasonal transitions, and their strength regarding process interactions, in combination with ordinary multivariate analysis techniques.

Finally, sensitivity tests for the model parameters are planned to quantify the effect of parameter sensitivity on the models' reproducibility.

4 Summary

This paper presented an overview of the GTMIP activity and the experiment protocol for the Stage 1 ~~of the GTMIP~~intercomparison, with site simulations using the GRENE-TEA site observation data in the pan-Arctic region for the previous three decades. We~~ have~~ described the framework of our project including targets, and provided datasets, conditions on model integration~~, analysis plans~~, lists of model output variables, and the habitat of currently participating models. We also included ~~the~~analysis plans and exemplary results ~~from a preliminary analysis~~ to give an outlook of the model–model and model–observation ~~comparison~~comparisons with respect to the major metrics defined for the energy budget, snowpack dynamics, and the carbon budget. ~~Through this~~This model intercomparison project ~~for~~was realized through a tight collaboration between the GRENE-TEA-participating modelling and field scientists. Additionally, we expect to offer insightful demonstrations of various cold-region terrestrial physical and biogeochemical ~~models, we will be able to offer insightful demonstrations of various~~ TPMs and valuable information for future improvements of the relevant models. All meteorological driving data for this project have already been made publicly available through ADS. The model outputs and comprehensive results from the GTMIP, which we hope will provide a useful benchmark dataset for the community, will also be available to the public at the end of the project.

Acknowledgements

This study is supported by the GRENE Arctic Climate Change Research Project, Ministry of the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

Abramowitz, G., Leuning, R., Clark, M., and Pitman, A.: Evaluating the performance of land surface models, J. Climate, 21, 5468-5481, 2008.

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Chang, A., Ferraro, R., Xie, P. P., Janowiak, J., Rudolf, B., Scheneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P., and Nelkin, E.: The Version-2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979–Present), J. Hydrometeor., 4, 1147-1167, 2003.

Ahlstrom, A., Schurgers, G., Arneth, A., and Smith, B.: Robustness and uncertainty in terrestrial ecosystem carbon response to CMIP5 climate change projections~~.~~, Environ. Res. Lett., 7, 044008, 2012.

Anav, A., Friedlingstein, P., Kidston, M., Bopp, L., Ciais, P., Cox, P., Jones, C., Jung, M., Myneni, R., and Zhu, Z.,: Evaluating the land and ocean components of the global carbon cycle in the CMIP5 Earth system models, J. Clim., 26, 6801-6843, 2013.

Arora, V. K., Boer, G. J., Friedlingstein, P., Eby, M., Jones, C. D., Christian, J. R., Bonan, G., Bopp, L., Brovkin, V., Cadule, P., Hajima, T., Ilyina, T., Lindsay, K., Tjiputra, J. F., and Wu, T.: Carbon-concentration and carbon-climate feedbacks in CMIP5 Earth system models~~.~~, J. Clim., 26, 5289-5314, 2013.

Brown, R. D. and Robinson, D. A.: Northern Hemisphere spring snow cover variability and change over 1922–2010 including an assessment of uncertainty~~.~~, Cryosphere, 5, 219-229, 2011.

Brutel-Vuilmet, C., Menegoz, M., and Krinner, G.: An analysis of present and future seasonal Northern Hemisphere land snow cover simulated by CMIP5 coupled climate models~~.~~, Cryosphere, 7, 67-80, 2013.

Cramer, W., Kicklighter, D. W., Bondeau, A., Moore, B., Churkina, G., Nemry, B., Ruimy, A., Schloss, A. L., and the participants of the Potsdam NPP model intercomparison.: Comparing global models of terrestrial net primary productivity (NPP): overview and key results~~.~~, Global change biol., 5 (S1), 1-15, 1999.

de Gonçalves, L. G. G., Borak, J. S., Costa, M. H., Saleska, S. R., Baker, I., Restrepo-Coupe, N., Muza, M. N., Poulter, B., Verbeeck, H., Fisher, J. B., Arain, M. A., Arkin, P., Cestaro, B. P., Christoffersen, B., Galbraith, D., Guan, X., van den Hurk, B. J. J. M., Ichii, K., Imbuzeiro,

H. M. A., Jain, A. K., Levine, N., Lu, C., Miguez-Macho, G., Roberti, D. R., Sahoo, A., Sakaguchi, K., Schaefer, K., Shi, M., Shuttleworth, W. J., Tian, H., Yang, Z.-L., and Zeng, X.: Overview of the large-scale biosphere–atmosphere experiment in Amazonia Data Model Intercomparison Project (LBA-DMIP), Agricultural and Forest Meteorology, 182–183, 111-127, 2013.

Dee, D. P., Uppalaa, S. M., Simmonsa, A. J. Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., H´olm, E. V., Isaksen, L., Kallberg, P., Kohler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. R. Meteorol. Soc., 137, 553-597, 2011.

Dirmeyer, P. A., Schlosser, C. A., and Brubaker, K. L.: Precipitation, recycling, and land memory: An integrated analysis, J. Hydrometeor., 10, 278-288, 2009.

Ekici A., Chadburn, S., Chaudhary, N., Hajdu, L. H., Marmy, A., Peng, S., Boike, J., Burke, E., Friend, A. D., Hauck, C., Krinner, G., Langer, M., Miller, P. A., and Beer, C.: Site-level model intercomparison of high latitude and high altitude soil thermal dynamics in tundra and barren landscapes, The Cryosphere Discuss., 8, 4959-5013, www.the-cryosphere-discuss.net/8/4959/2014/, doi:10.5194/tcd-8-4959-2014, 2014.

Etchevers, P., E. Martin, E., Brown, R., Fierz, C., Lejeune, Y., Bazile, E., Boone, A., Dai, Y. J., Essery, R., Fernandez, A., Gusev, Y., Jordan, R., Koren, V., Kowalczyk, E., Nasonova, N. O., Pyles, R. D., Schlosser, A., Shmakin, A. B., Smirnova, T. G., Strasser, U., Verseghy, D., Yamazaki, T., Yang, Z. L.: Validation of the surface energy budget simulated by several snow models~~.~~, Ann. Glaciol., 38, 150-158, 2004.

Essery, R., Rutter, N., Pomeroy, J., Baxter R., Stahli, M., Gustafsson, D., Barr, A., Bartlett, P., and Elder, K.: SnowMIP2: An evaluation of forest snow process simulations~~.~~, Bulletin of the American Meteorological Society, 90, 1120-1135, doi:10.1175/2009BAMS2629.1, 2009.

Friedlingstein, P., Cox, P., Betts, R., Bopp, L., Bloh, W. V., Brovkin, V., Cadule, P., Doney, S., Eby, M., Fung, I., Bala, G., John, J., Jones, C., Joos, F., Kato, T., Kawamiya, M., Knorr, W., Lindsay, K., Matthews, H. D., Raddatz, T., Rayner, P., Reick, C., Roeckner, E., Schnitzler, K. G., Schnur, R., Strassmann, K., Weaver, J., Yoshikawa, C., and Zeng, N.:

Climate–carbon cycle feedback analysis: Results from the C 4 MIP model intercomparison~~.~~, Journal of Climate 19, 3337-3353, 2006.

Hajima, T., Tachiiri, K., Ito, A., and Kawamiya, M.: Uncertainty of concentration–terrestrial carbon feedback in Earth System Models, Journal of Climate, 27, 3425-3445, 2014.

Henderson-Sellers, A., Yang, Z. L., and Dickinson, R. E.: The Project for intercomparison of land surface schemes (PILPS), Bull. Amer. Meteor. Soc., 74, 1335-1349, 1993.

Harris, I., Jones, P. D., Osborn, T. J., and Listera, D. H.: Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset, Int. J. Climatology, 34, 623-642, 2014.

~~Hirashima, H., Ohata, T., Kodama, Y., Yabuki, H., Sato, N., and Geoggiadi, A.: Nonuniform Distribution of Tundra Snow Cover in Eastern Siberia, J. Hydrometeor., 5, 373-389, 2004.~~

Huntzinger, D. N., Schwalm, C., Michalak, A. M., Schaefer, K., King, A. W., Wei, Y., Jacobson, A., Liu, S., Cook, R. B., Post, W. M., Berthier, G., Hayes, D., Huang, M., Ito, A., Lei, H., Lu, C., Mao, J., Peng, C. H., Peng, S., Poulter, B., Riccuito, D., Shi, X., Tian, H., Wang, W., Zeng, N., Zhao, F., and Zhu, Q.: The North American carbon program multi-scale synthesis and terrestrial model intercomparison project – Part 1: Overview and experimental design, Geosci. Model Dev., 6, 2121–2133, doi:10.5194/gmd-6-2121-2013, 2013.

Ichii, K., Kondo, M., Lee, Y.-H., Wang, S.-Q., Kim, J., Ueyama, M., Lim, H.-J., Shi, H., Suzuki, T., Ito, A., Kwon, H., Ju, W., Huang, M., Sasai, T., Asanuma, J., Han, S., Hirano, T., Hirata, R., Kato, T., Li, S.-G., Li, Y.-N., Maeda, T., Miyata, A., Matsuura, Y., Murayama, S., Nakai, Y., Ohta, T., Saitoh, T., Saigusa, N., Takagi, K., Tang, Y.-H., Wang, H.-M., Yu, G.-R., Zhang, Y.-P., and Zhao, F.-H.: Site-level model–data synthesis of terrestrial carbon fluxes in the CarboEastAsia eddy-covariance observation network: toward future modeling efforts, Journal of Forest Research, 18, 13-20, 2013.

Ichii, K., Suzuki, T., Kato, T., Ito, A., Hajima, T., Ueyama, M., Sasai, T., Hirata, R., Saigusa, N., Ohtani, Y., and Takagi, K.: Multi-model analysis of terrestrial carbon cycles in Japan: limitations and implications of model calibration using eddy flux observations, Biogeosciences, 7, 2061-2080, 2010.

Iwahana, G., Takano, S., Petrov, R. E., Tei, S., Shingubara, R., Maximov, T. C., Fedrov, A. N., Desyatkin, A. R., Nikolaev, A. N., Desyatkin, R. V., Sugimoto, A.: Geocryological characteristics of the upper permafrost in a tundraforest transition of the Indigirka River Valley, Russia, Polar Science, http://dx.doi.org/10.1016/j.polar.2014.01.005, 2014.

Jung, M., Le Maire, G., Zaehle, S., Luyssaert, S., Vetter, M., Churkina, G., Ciais, P., Viovy, N., and Reichstein, M.: Assessing the ability of three land ecosystem models to simulate gross carbon uptake of forests from boreal to Mediterranean climate in Europe, Biogeosciences, 4, 647-656, 2007.

Kodama, Y., Sato, N., Yabuki, H., Ishii, Y., Nomura, M., and Ohata, T.: Wind direction dependency of water and energy fluxes and synoptic conditions over a tundra near Tiksi, Siberia, Hydrological Processes, 21, 2028-2037, 2007.

Kotani, A., Kononov, A. V., Ohta, T., and Maximov, T. C.: Temporal variations in the linkage between the net ecosystem exchange of water vapour and CO_2 over boreal forests in eastern Siberia, Ecohydrology, DOI: 10.1002/eco.1449, 2013.

Koven, C. D., Ringeval, B., Friedlingstein, P., Ciais, P., Cadule, P., Khvorostyanov, D., Krinner, G., and Tarnocai, C.: Permafrost carbon-climate feedbacks accelerate global warming~~.~~, Proc. Natl. Acad. Sci. USA, 108, 14 769–14 774, doi:10.1073/pnas.1103910108, 2011.

Koven, C. D., Riley W. J., and Stern, A.: Analysis of permafrost thermal dynamics and response to climate change in the CMIP5 Earth System Models~~.~~, J. Clim., 26, 1877–1900, 2013.

Lopez M.L., Saito H, Kobayashi K, Shirota T, Iwahana G, Maximov T. C., and Fukuda M.: Interannual environmental-soil thawing rate variation and its control on transpiration from Larix cajanderi, Central Yakutia, Eastern Siberia~~.~~, Journal of Hydrology 338: 251–260. DOI: 10.1016/j. jhydrol.2007.02.039, 2007.

Morales, P., Sykes, M. T., Prentice, I. C., Smith, P., Smith, B., Bugmann, H., Zierl, B., Friedlingstein, P., Viovy, N., Sabate, S., Sanchez, A., Pla, E., Gracia, C. A., Sitch, S., Arneth, A., and Ogee, J.: Comparing and evaluating process-based ecosystem model predictions of carbon and water fluxes in major European forest biomes, Global Change Biology, 11, 2211-2233, 2005.

McGuire, A. D., Christensen, T. R., Hayes, D., Heroult, A., Euskirchen, E., Kimball, J. S., Koven, C., Lafleur, P., Miller, P. A., Oechel, W., Peylin, P., Williams, M., and Yi, Y.: An assessment of the carbon balance of Arctic tundra: comparisons among observations, process models, and atmospheric inversions, Biogeosciences, 9, 3185-3204, 2012.

Nakai, Y., Matsuura, Y., Kajimoto, T., Abaimov, A. P., Yamamoto, S., and Zyryanova, O. A.: Eddy covariance CO_2 flux above a Gmelin larch forest in continuous permafrost of central Siberia during a growing season~~.~~, Theor. Appl. Climatol. 93, 133-147, http://dx.doi.org/10.1007/s00704-007-0337-x, 2008.

Nakai, T., Kim, Y., Busey, R. C., Suzuki, R., Nagai, S., Kobayashi, H., Park, H., Sugiura, K., and Ito, A.: Characteristics of evapotranspiration from a permafrost black spruce forest in interior Alaska, Polar Science, 7, 136-148, 2013.

Ohta, T., Hiyama, T., Tanaka, H., Kuwada, T., Maximov, T. C., Ohata, T., Fukushima, Y.: Seasonal variation in the energy and water exchanges above and below a larch forest in Eastern Siberia, Hydrol. Process. 15, 1459-1476, 2001.

Ohta, T., Maximov, T. C., Dolman, A. J., Nakai, T., van der Molen, M. K., Kononov, A. V., Maximov, A. P., Hiyama, T., Iijima, Y., Moors, E. J., Tanaka, H., Toba, T., Yabuki, H.: Interannual variation of water balance and summer evapotranspiration in an Eastern Siberian larch forest over a 7-year period (1998–2006), Agric. Forest Meteorol. 148, 1941-1953, 2008.

Ohta, T., Kotani, A., Iijima, Y., Maximov, T. C., Ito, S., Hanamuraa, M., Kononov, A. V., and Maximov, A. P.: Effects of waterlogging on water and carbon dioxide fluxes and environmental variables in a Siberian larch forest, 1998–2011, Agricultural and Forest Meteorology, 188, 64-75, 2014.

Piao, S., Stich, S., Ciais, P., Friedlingstein, P., Peylin, P., Wang, X., Ahstrom, A., Anav, A., Candell, J. G., Cong, N., Huntingford, C., Jung, M., Levis, S., Levy, P. E., Li, J., Lin, X., Lomas, M. R., Lu, M., Luo, Y., Ma, Y., Myneni, R. B., Poulter, B., Sun, Z., Wang, T., Viovy, N., Zaehle, S., and Zeng, N.: Evaluation of terrestrial carbon cycle models for their response to climate variability and CO_2 trends, Global Change Biol., 19, 2117-2132, 2013.

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, Glob. Change Biol. 16, 641-56, 2010.

Rawlins, M. A., McGuire, A. D., ,Kimball, J. K., Dass, P., Lawrence, D., Burke, E., Chen, X., Delire, C., Koven, C., MacDougall, A., Peng, S., Rinke, A., Saito, K., Zhang, W., Alkama, R., Bohn, T. J., Ciais, P., Decharme, B., Gouttevin, I., Hajima, T., Ji, D., Krinner, G., Lettenmaier,

D. P., Miller, P., Moore, J. C., Smith, B., and Sueyoshi, T.: Assessment of model estimates of land–atmosphere CO_2 exchange across Northern Eurasia, Biogeosciences Discuss., 12, 2257-2305, doi:10.5194/bgd-12-2257-2015, 2015.

Saito, K., Yasunari, T., and Takata, K.: Relative roles of large-scale orography and land surface processes in the global hydroclimate. Part II: Impacts on hydroclimate over Eurasia, J. Hydrometeor., 7, 642-659, 2006.

Saito, K. Miyazaki, S., Mori, J., Ise, T., Arakida, H., Sueyoshi, T., Hajima, T., Iijima, Y., Yamazaki, T., and Sugimoto, A.: GTMIP meteorological driving dataset for the GRENE-TEA observation sites (level 0.2), 0.20, Arctic Data archive System (ADS), Japan, https://ads.nipr.ac.jp/dataset/A20141009-005, 2014a.

Saito, K., Miyazaki, S., Mori, J., Ise, T., Arakida, H., Suzuki, R., Sato, A., Iijima, Y., Yabuki, H., Iijima, Y., Sueyoshi, T., Hajima, T., Sato, H., Yamazaki, T., and Sugimoto, A.: GTMIP meteorological driving dataset for the GRENE-TEA observation sites (level 1.0), 1.00, Arctic Data archive System (ADS), Japan, https://ads.nipr.ac.jp/dataset/A20141009-006, 2014b.

Saito, K., Sasai, T., Miyazaki, S., Mori, J., Ise, T., Arakida, H., Sueyoshi, T., Hajima, T., Iijima, Y., Yamazaki, T., and Sugimoto, A.: GTMIP fraction of photosynthetically active radiation (fPAR) and leaf area index (LAI) for the GRENE-TEA observation sites (level 1.0), 1.00, Arctic Data archive System (ADS), Japan, https://ads.nipr.ac.jp/dataset/A20141009-007, 2014c.

Saito, K., Miyazaki, S., Mori, J., Ise, T., Arakida, H., Suzuki, R., Sato, A., Iijima, Y., Yabuki, H., Iijima, Y., Sueyoshi, T., Hajima, T., Sato, H., Yamazaki, T., Sugimoto, A.: GTMIP meteorological driving dataset for the GRENE-TEA observation sites (20-year detrended), 1.00, Arctic Data archive System (ADS), Japan, https://ads.nipr.ac.jp/dataset/A20150205-001, 2015.

Sasai, T., Saigusa, N., Nasahara, K. N., Ito, A., Hashimoto, H., Nemani, R. R., Hirata, R., Ichii, K., Takagi, K., Saitoh, T. M., Ohta, T., Murakami, K., Yamaguchi, Y., and Oikawa, T.: Satellite-driven estimation of terrestrial carbon flux over Far East Asia with 1-km grid resolution, Remote Sensing of Environment, 115, 7, 1758-1771, doi:10.1016/j.rse.2011.03.007., 2011.

Sato, A., Kubota, H., Matsuda, M., and Sugiura, K.: Seasonal variation of heat exchange in the boreal forest of Finnish Lapland, Second Wadati Conference, on Global Change and the Polar Climate, extended abstracts, 228-230, 2001.

Schwalm, C. R., Williams, C. A., Schaefer, K., Anderson, R., Arain, M. A., Baker, I., Barr, A., Black, T. A., Chen, G., Chen, J. M., Ciais, P., Davis, K. J., Desai, A., Dietze, M., Dragoni, D., Fischer, M. L., Flanagan, L. B., Grant, R., Gu, L., Hollinger, D., Izaurralde, R. C., Kucharik, C., Lafleur, P., Law, B. E., Li, L., Li, Z., Liu, S., Lokupitiya, E., Luo, Y., Ma, S., Margolis, H., Matamala, R., McCaughey, H., Monson, R. K., Oechel, W. C., Peng, C., Poulter, B., Price, D. T., Riciutto, D. M., Riley, W., Sahoo, A. K., Sprintsin, M., Sun, J., Tian, H., Tonitto, C., Verbeeck, H., and Verma, S. B.: A model-data intercomparison of CO_2 exchange across North America: Results from the North American Carbon Program site synthesis, Journal of Geophysical Research, 115, 2010.

Slater, A. G. and Lawrence, D. M.: Diagnosing present and future permafrost from climate models~~.~~, J. Clim., doi:10.1175/JCLI-D-12-00341.1, 2013.

Sueyoshi, T., Saito, K., Miyazaki, S., Mori, J., Ise, T., Arakida, H., Suzuki, R., Sato, A., Iijima, Y., Yabuki,H., Hajima, T., Sato, H., Yamazaki, T., Sugimoto, A.,:GRENE-TEA Model Intercomparison Project (GTMIP) forcing and evaluation dataset, Earth System Science Data, ~~to be submitted~~accepted for publication as a discussion paper in ESSDD, 2015.

Takata, K.: Sensitivity of land surface processes to frozen soil permeability and surface water storage, Hydrological Processes, 16, 2155-2172, 2002.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, Bull. Amer. Meteor. Soc., 93, 485-498, 2012.

~~Wang, W., Rinke,~~Todd-Brown, K. E. O., Randerson, J. T., Hopkins, F., Arora, V., Hajima, T., Jones, C., Shevliakova, E., Tjiputra, J., Volodin, E., Wu, T., Zhang, Q., and Allison, S. D.: Changes in soil organic carbon storage predicted by Earth system models during the 21st century, Biogeosciences, 11, 2341-2356, doi:10.5194/bg-11-2341-2014, 2014.

~~A., Moore, J. C., Cui, X., Ji, D., Li, Q., Zhang, N., Wang, C., Zhang, S., Lawrence, D. M., McGuire, A. D., Zhang, W., Delire, C., Koven, C., Saito, K., MacDougall, A. Burke, E., and Decharme, B.: Diagnostic and model dependent uncertainty of simulated Tibetan permafrost area, The Cryosphere Discuss., 9, 1769-1810, 2015~~

Watanabe K, Mizoguchi, M., Kiyosawa, H., and Kodama Y.: Properties and horizons of active layer soils in tundra at Tiksi, Siberia~~.~~, Journal of Japan Society of Hydrology and Water Resources 13(1), 9-16, (in Japanese with English abstract), 2000.

Zaehle, S., Medlyn, B. E., De Kauwe, M. G., Walker, A. P., Dietze, M. C., Hickler, T., Luo, Y., Wang, Y.-P., El-Masri, B., Thornton, P., Jain, A., Wang, S., Warlind, D., Weng, E., Parton, W., Iversen, C. M., Gallet-Budynek, A., McCarthy, H., Finzi, A., Hanson, P. J., Prentice, I. C., Oren, R. and Norby, R. J.: Evaluation of 11 terrestrial carbon–nitrogen cycle models against observations from two temperate Free-Air CO_2 Enrichment studies, New Phytologist 202, 803–822. doi: 10.1111/nph.12697, 2014.

Zhang, T. J., Frauenfeld, O. W., Serreze, M. C., Etringer, A., Oelke, C., McCreight, J., Barry, R. G., Gilichinsky, D., Yang, D., Ye, H., Ling, F., and Chudinova, S.: Spatial and temporal variability in active layer thickness over the Russian Arctic drainage basin, J. Geophys. Res. Atmos., 110, D16101, 2005.

Table 1. The key process categories and target processes

A: Key processes categories	B: Target processes and metrics
Energy and water budget	Partition of energy and water at surface, canopy, and subsurface, albedo
Snowpack (snow cover ratio, snow depth/snow water equivalent)	Snow water equivalent, snow density, snow cover duration (length and dates)
Phenology	Annual maximum leaf area index, growing season (length and dates)
Ground freezing/thawing, active layer	Active layer thickness (in permafrost) or maximum seasonal frozen depth, trumpet curve, ice content ratio
Carbon budget	Net primary production, heterotrophic and autotrophic respiration, net ecosystem production, stored carbon mass in different pools, turnover rates

Table 2. The location, dominant vegetation type, soil, climate, fraction of photosynthetically active radiation (fPAR), possible data for validation, and references for observed data for (a) Fairbaks, (b) Kevo, (c) Tiksi, (d) Yakutsk, (e) Chokurdakh, and (f) Tura.

(a): Fairbanks (Poker Flat Research Range), Alaska, USA

Location	65°07’24” N, 147°29’15.” W
Altitude	210 m
Dominant vegetation type	Black spruce forest
Soil	0-14cm layer: moss 14-25cm: undecomposed organic layer 25-39cm: decomposed organic layer 39cm- : silt soil Active layer thickness: 43cm in 2013
Climate	Mean annual air temperature: -2.8 °C (2011) Annual precipitation: 312 mm (2011)
fPAR and LAI [1)]	fPAR: 0.03 (Jan), 0.05 (Feb), 0.05 (Mar), 0.13 (Apr), 0.39 (May), 0.69 (Jun), 0.69 (Jul), 0.69 (Aug), 0.43 (Sep), 0.23 (Oct), 0.06 (Nov), 0.00 (Dec) LAI: 0.05 (Jan), 0.09 (Feb), 0.09 (Mar), 0.23 (Apr), 0.99 (May), 2.26 (Jun), 2.32 (Jul), 1.90 (Aug), 0.80 (Sep), 0.49 (Oct), 0.10 (Nov), 0.01 (Dec.)
Data available for model validation	Snow depth, ground temperature (-0.05, -0.1, -0.2, -0.4, -1.0m), soil moisture (-0.05, -0.1, -0.2, -0.4m), leaf area index, albedo, FPAR (Fraction of photosynthetically active radiation), upward -short and long wave radiation, energy and carbon fluxes
Reference	Nakai et al., 2013

(b): Kevo (Kevo Research Station), Finland

Location	69°45' 25"N, 27°00' 37"E
Altitude	100m
Dominant vegetation type	Pine forest
Soil	0-20cm: humus soil
	20－50cm: sandy silt
Climate	Mean annual air temperature: -1.6 °C
	Annual precipitation: 415 mm
fPAR and LAI [1)]	fPAR: 0.03 (Jan), 0.06 (Feb), 0.08 (Mar), 0.11 (Apr), 0.51 (May), 0.56 (Jun), 0.69 (Jul), 0.76 (Aug), 0.68 (Sep), 0.45 (Oct), 0.10 (Nov), 0.02 (Dec)
	LAI: 0.05 (Jan), 0.10 (Feb), 0.14 (Mar), 0.21 (Apr), 1.13 (May), 1.63 (Jun), 2.52 (Jul), 2.78 (Aug), 1.66 (Sep), 1.18 (Oct), 0.21 (Nov), 0.05 (Dec.)
Data available for model validation	Snow depth, snow (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7m) and ground temperature (-0.1, -0.2, -0.3, -0.35m), soil moisture (-0.1, -0.2, -0.3m), albedo, upward short and long wave radiation
Reference	Sato et al., 2001

(c): Tiksi, Sakha Republic, Russian Federation

Location	71°35’21”N, 128°46’27”E
Altitude	40 m
Dominant vegetation type	Non-tussock sedge, dwarf-shrubs, and moss tundra
Soil	0-1cm: partially decomposed litter 1-15cm: loam 15-70cm: silt with gravel Active layer thickness: 70cm
Climate	Mean annual air temperature: -13.5 °C Annual precipitation: 331 mm
fPAR and LAI 1)	fPAR: 0.00 (Jan), 0.00 (Feb), 0.00 (Mar), 0.00 (Apr), 0.03 (May), 0.29 (Jun), 0.45 (Jul), 0.47 (Aug), 0.28 (Sep), 0.04 (Oct), 0.00 (Nov), 0.00 (Dec) LAI: 0.00 (Jan), 0.00 (Feb), 0.00 (Mar), 0.00 (Apr), 0.05 (May), 0.52 (Jun), 0.88 (Jul), 0.73 (Aug), 0.49 (Sep), 0.07 (Oct), 0.00 (Nov), 0.00 (Dec.)
Data available for model validation	Snow depth, ground temperature (-0.1, -0.2, -0.3, -0.47, -1, -2, -3, -5, -10, -20, -30m), soil moisture (0, -0.05, -0.15, -0.3m), albedo, upward short and long-wave radiation
Reference	Kodama et al., 2007; Watanabe et al., 2000

(d): Yakutsk (Spasskaya Pad), Sakha Republic, Russian Federation

Location	62°15’18”N, 129°37’6”E
Altitude	220 m
Dominant vegetation type	Larch forest
Soil	0-20cm: organic layer Upper mineral layer: sandy loam Lower mineral layer: silty loam (More than 80% of root: within a soil depth of 20 cm) Active layer thickness: 1.2m
Climate	Mean annual air temperature: -10.2 °C Annual precipitation: 188 mm
fPAR and LAI [1)]	fPAR: 0.00 (Jan), 0.00 (Feb), 0.00 (Mar), 0.05 (Apr), 0.28 (May), 0.46 (Jun), 0.42 (Jul), 0.21 (Aug), 0.03 (Sep), 0.00 (Oct), 0.00 (Nov), 0.02 (Dec) 0.00 LAI: 0.00 (Jan), 0.00 (Feb), 0.00 (Mar), 0.00 (Apr), 0.07 (May), 0.58 (Jun), 1.05 (Jul), 0.81 (Aug), 0.28 (Sep), 0.04 (Oct), 0.00 (Nov), 0.00 (Dec.)
Possible data for model validation	Snow depth, ground temperature (-0.1, -0.2, -0.4, -0.6, -0.8, -1.2), soil moisture (-0.1, -02, -0.4, -0.6, -0.8m), albedo, FPAR, upward short and long wave radiation, energy and carbon fluxes
Reference	Ohta et al., 2001, 2008, 2014; Kotani et al., 2013; Lopez et al., 2007

(e): Chokurdakh (Kodack/Krybaya~~) ,~~), Sakha Republic, Russian Federation

Location	70°33’48”N, 148°15’51”E
Altitude	9 m
Dominant vegetation type	Tussock wetland/shrubs/sparse larch trees
Soil	Clay loam, silty clay loam Active layer thickness: 0.4-0.7m
Climate	Mean annual air temperature: -13.4 °C Annual precipitation: 196 mm
fPAR and LAI 1)	fPAR: 0.00 (Jan), 0.00 (Feb), 0.00 (Mar), 0.00 (Apr), 0.00 (May), 0.01 (Jun), 0.18 (Jul), 0.45 (Aug), 0.48 (Sep), 0.26 (Oct), 0.07 (Nov), 0.02 (Dec) LAI: 0.00 (Jan), 0.00 (Feb), 0.00 (Mar), 0.00 (Apr), 0.02 (May), 0.32 (Jun), 0.91 (Jul), 0.79 (Aug), 0.41 (Sep), 0.15 (Oct), 0.00 (Nov), 0.00 (Dec.)
Data available for model validation	Ground temperature (-0.01, -0.05, -0.1, -0.2, -0.3, -0.4, -0.5, -0.75, -1.0, -1.5, -2.0, -2.5, -3.0, -4.0, -5.0, -5.5, -7.0, -10.0 m), soil moisture (-0.035, -0.145, -0.335, -0.535m), albedo, upward short and long-wave radiation, energy and carbon fluxes
Reference	Iwahana et al., 2014

(f): Tura, Russian Federation

Location	64°12’32”N, 100°27’49”E
Altitude	250 m
Dominant vegetation type	Larch forest (average age: 105 years in 2005)
Soil	10-20cm organic layer Cryosol Active layer thickness: 1m
Climate	Mean annual air temperature: -8.9 °C Annual precipitation: 360 mm
fPAR and LAI average value extracted from 1km grid MODIS satellite from 2001 to 2011 (Sasai et al., 2011)	fPAR: 0.00 (Jan), 0.00 (Feb), 0.00 (Mar), 0.01 (Apr), 0.20 (May), 0.48 (Jun), 0.52 (Jul), 0.49 (Aug), 0.29 (Sep), 0.10 (Oct), 0.00 (Nov), 0.00 (Dec) LAI: 0.00 (Jan), 0.00 (Feb), 0.00 (Mar), 0.01 (Apr), 0.46 (May), 1.28 (Jun), 1.43 (Jul), 1.17 (Aug), 0.48 (Sep), 0.17 (Oct), 0.00 (Nov), 0.00 (Dec.)
Data available for model validation	Ground temperature (-0.05, -0.1, -0.2, -0.4, -0.5), soil moisture (-0.05, -0.1, -0.2, -0.4, -0.5), albedo, FPAR, upward short and long-wave radiation, energy and carbon fluxes
Reference	Nakai et al., 2008

1) Average values extracted from 1 km grid MODIS satellite from 2001 to 2011 (Sasai et al., 2011)

Table 3. The list of metrics for model performance evaluation for (a) energy and water budgets, (b) snowpack, (c) phenology, (d) subsurface hydrological and thermal states, and (e) the carbon budget.

(a): Energy and water budget

Variable	Definition	Units	Direction (+)	Time step
Rn_season, Rn_annual	Seasonally and annually averaged net radiation	W/m^2	Downward	seasonal annual
Qh_season, Qh_annual	Seasonally and annually averaged sensible heat flux	W/m^2	Upward	seasonal annual
Qle_season, Qle_annual	Seasonally and annually averaged latent heat flux	W/m^2	Upward	seasonal annual
ET_season, ET_annual	Seasonally and annually averaged total evapotranspiration	mm/day	Upward	seasonal annual
Qs_season, Qs_annual	Seasonally and annually averaged surface runoff	mm/day	Out of soil column	seasonal annual
Qsb_season, Qsb_annual	Seasonally and annually averaged subsurface runoff	mm/day	Out of soil column	seasonal annual
Et_veg_season, Et_veg_annual	Seasonally and annually averaged transpiration of vegetation	mm/day	Upward	seasonal annual
E_soil_season, E_soil_annual	Seasonally and annually averaged soil evaporation	mm/day	Upward	seasonal annual
Wg_frac_season Wg_frac_annual	Seasonally and annually averaged fraction of saturation of soil water content (wilting=0, saturation=1)	-	-	seasonal annual
deltaWg_season, deltaWg_annual	Seasonally and annually averaged change of stored soil moisture	mm/day	-	seasonal annual
alpha_season, alpha_annual	Seasonally and annually averaged shortwave albedo	-	-	seasonal annual
E_can_season, E_can_annual	Seasonally and annually averaged canopy interception evaporation	mm/day	Upward	seasonal annual

(b): Snowpack

Variable	Definition	Units	Direction (+)	Time step
SWE_max Date_SWE_max	Annual maximum snow water equivalent and the date reached	kg/m^2 day	-	annual
SnD_max Date_SnD_max	Annual maximum snow depth and the date reached	m day	-	annual
SnowDuration Date_start_snow_cover	Annual duration of snow cover h and the date of snow cover start/end	day	-	annual
Sub_snow_season, Sub_snow_annual	Seasonally and annually averaged total sublimation from the ground snow pack	mm/day	Upward	annual

(c): Phenology

Variable	Definition	Units	Direction (+)	Time step
LAI_max	Annual maximum leaf area index	m^2/m^2	-	annual
GrowSeasonLentgh	Growing season length and the date of start/end of growing season	day	-	annual

(d): Subsurface hydrological and thermal states

Variable	Definition	Units	Direction (+)	Time step
ALT or ThawDepth_max	Active layer thickness (permafrost region) or annual maximum thawing depth (seasonal frozen ground) and the date reached	m	-	annual
FrozenDepth_max	Annual maximum frozen depth and the date reached	m	-	annual
Tg_range_*depth*	Annual range of soil temperature in pre-defined soil layer	K	-	annual
Wg_frozfrac_max_*depth*	Annual maximum fraction of soil moisture mass in the solid phase in pre-defined soil layer	-	-	annual

(e): Carbon budget

Variable	Definition	Units	Direction (+)	Time step
NPP_annual NPP_growing	Annual and growing season net primary production on land	kgC/m^2/year kgC/ m^2/duration	Downward	annual growing season
GPP_annual GPP_growing	Annual gross primary production	kgC/m^2/year kgC/ m^2/duration	Downward	annual growing season
Rh_annual Rh_growing	Annual heterotrophic respiration on land	kgC/m^2/year kgC/ m^2/duration	Upward	annual growing season
Ra_annual Ra_growing	Annual autotrophic (plant) respiration on land	kgC/m^2/year kgC/ m^2/duration	Upward	annual growing season
NEP_annual NEP_growing	Annual net ecosystem productivity (=NPP-Rh) on land	kgC/m^2/year kgC/ m^2/duration	Downward	annual growing season
Re_annual Re_growing	Annual and growing season ecosystem respiration (=Ra + Rh) on land	kgC/m^2/year kgC/ m^2/duration	Downward	annual growing season
cBiomass_annual	Stored carbon mass in biomass pool	kgC/m^2	-	annual
TotCarLitSoil	Stored carbon mass in litter pool and soil	kgC/m^2	-	annual
cTurnoverRate_biomass	Turnover rate of carbon in biomass pool	1/year	-	-
cTurnoverRate_soil	Turnover rate of carbon litter pool and soil	1/year	-	-

Figure Captions

Figure 1. ~~Location map~~“Pirates of the ~~GRENE-TEA sites~~Arctic” sit at the Round Table

Figure 2. Schematic diagram for stages 1 and 2 of GTMIP

Figure 3. Location map of the GRENE-TEA sites

Figure 4. The habitat of models participating in the GTMIP. The vertical and horizontal axes show the ratio of the incorporation of biogeochemical processes and physical processes, respectively.

Figure ~~3. Comparison~~5. Example comparison of model outputs with observations, and the inter-model range for the annual mean latent heat flux for averages from 1980 to 2013. The results of biogeochemical and physical models are shown ~~the~~by boxes and lines in orange and blue, respectively. The biogeochemical models ~~include~~included are BEAMS, Biome-BGC, CHANGE, SEIB-DGVM, and VISIT~~. The~~, while the physical models ~~include~~are 2LM, JULES, MATSIRO, and PB-SDM. The orange and blue horizontal lines indicate medians. The bottom and top of the boxes correspond to the 25th and 75th percentiles of the average values, for 1980 to 2013 (except BEMAS, which is for 2001 to 2011), of model outputs. The bottom and top of the lines show the minimum and maximum outputs from the participating models, respectively. The dots show the observed average values for 2011, 2012, and 2013 at FB and for 1998, 2001, 2003, 2004, 2007, and 2008 at YK.

Figure ~~4~~6. As for Fig. 3, except the plot displays annual maximum snow depth. The physical models include 2LM, JULES, MATSIRO, PB-SDM, SMAP, and SNOWPACK (for FB and KVTK only). The observation shows the average values for 1980–2012, 1996–2013, 1980–2008, and 1980–2008 at FB, KV, TK, and YK, respectively.

Figure ~~5~~7. As for Fig. 3, except the plot displays annual gross primary production. The relevant biogeochemical models include BEAMS, Biome-BGC, CHANGE, LPJ, SEIB-DGVM, STEM1, and VISIT. The observation shows the average values for 2011–2013 and 2004–2012 at FB and YK, respectively.

Figure ~~6~~8. As for Fig. 5, except the plot displays annual net primary production.

Figure 9. Example of seasonal transitions in ground temperature, snow, and vegetation among models.