Summary of Research (Final Technical Report)

Investigating the Role of Biogeochemical Processes in the Northern High Latitudes on Global Climate Feedbacks Using an Efficient Scalable Earth System Model

A Research Project Funded by the Office of Science (BER), US DOE, Award No: DOE-DE-SC0006706

Principal Investigator

Atul K Jain
Department of Atmospheric Science
105 S. Gregory Street, Urbana, IL 61801

Phone: 217-333-2128
Email: jain1@illinois.edu

Overall Objectives and Accomplishments

The overall objectives of this DOE funded project is to combine scientific and computational challenges in climate modeling by expanding our understanding of the biogeophysical-biogeochemical processes and their interactions in the northern high latitudes (NHLs) using an earth system modeling (ESM) approach, and by adopting an adaptive parallel runtime system in an ESM to achieve efficient and scalable climate simulations through improved load balancing algorithms.

As part of this research we have accomplished most of the tasks that we originally proposed, leading to the completion of the proposal. Under this grant, 27 journal articles and 21 conference proceeding articles were published. All these articles acknowledged financial support of this grant. In addition, this grant supported one Ph.D. student. The student completed his Ph.D. research in August 2014.

In the following, the significant findings from the DOE supported study are outlined below. The findings are described in two integrated areas: (1) Extension of the capabilities of the ISAM land surface model capability to evaluate key interactions amongst Earth's climate and terrestrial processes in the NHL regions and (2) applying various novel computational methods to ISAM land surface model to achieve dynamic and improved modeling scalability.

(1) Extension of ISAM Land Surface Model for NHLs

Given the current challenges in modeling the northern high-latitudes (NHLs) regions above permafrost and their growing importance in global climate studies, here we placed emphasis on studying the biogeophysical-biogeochemical interactions in these regions. First, we applied the model from flux tower sites to the global scales to investigate the impacts of environmental uncertainties on modeled terrestrial carbon, energy and water fluxes (Section 1.1 and 1.2). in keeping with our site-level to global scale analysis, we also investigated the relative importance of these processes in comparison with modeling uncertainties from meteorological datasets (Section 1.1 and 1.2). Next, we focused on the northern high-latitude regions (NHLs) (45—90^{o}N) to address contemporary issues related to vegetation and soil biogeophysical-biogeochemical interactions. In regards to vegetation for NHL regions, we implemented NHL ecosystem-specific dynamic phenology and dynamic rooting distribution and depth parameterizations (Section 1.3). In regards to soil and snow processes in NHLs, we extend the permafrost biogeophysics in ISAM, by incorporating recent advances in key soil/snow processes that affect permafrost dynamics; analysis of influence of these processes on permafrost area and stability in the NHLs (Section 1.4). Next, we investigated the response of modeled permafrost soil organic carbon stocks to the improved representation of these soil/snow biogeophysical processes (Section 1.5). As part of this project, we also coupled the current version of ISAM with the Community ESM (CESM1), to develop a flexible ESM modeling framework: CESM-ISAM (Section 1.6). We also applied the well calibrated and tested model for a number of studies, model-model and data-model intercomparison studies (Section 1.7).

The specific model utilized here is the Integrated Science Assessment Model (ISAM) land surface model, which estimates terrestrial energy, water, momentum, and carbon fluxes at hourly timescales (Figure 1).

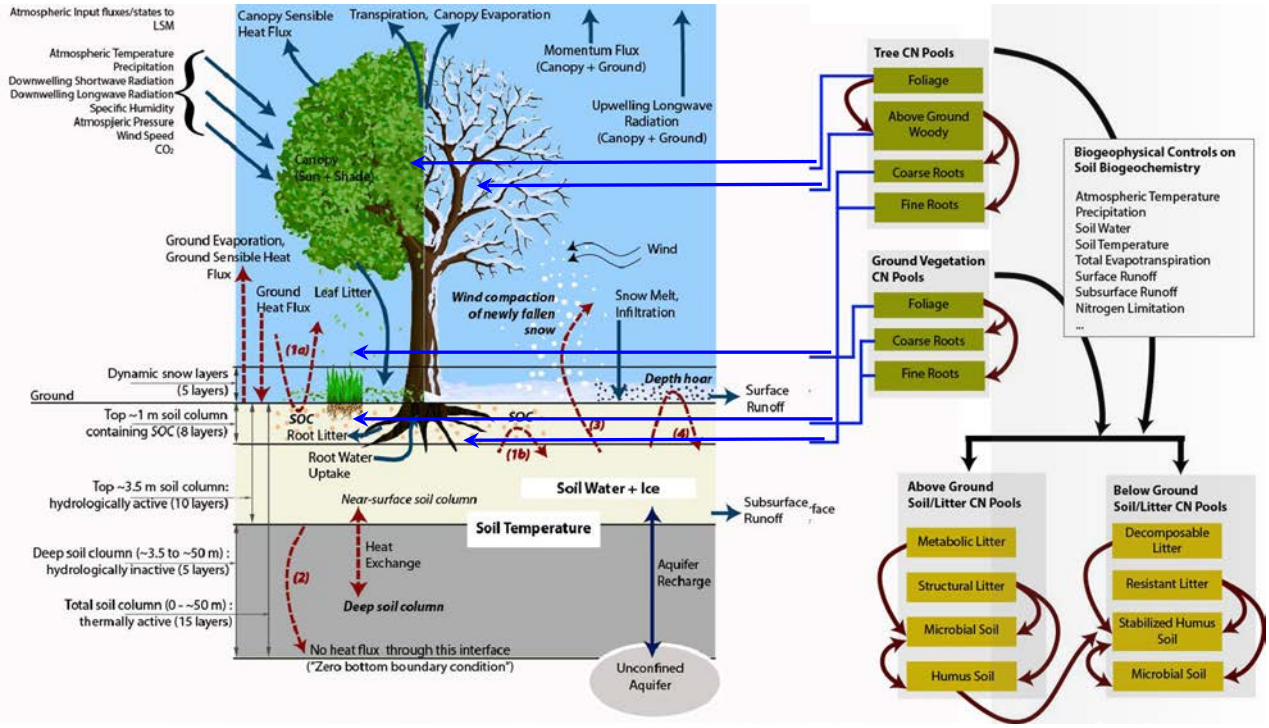


Figure 1.1: Conceptual diagram of coupled biogeophysics and biogeochemistry ISAM land model, focusing on the four key cold-region soil/snow processes recently implemented is ISAM (Barman and Jain, 2016). Specifically, they are: effects of *SOC* on thermal and hydrological properties, incorporation of a deep soil column, wind compaction of snow, and depth hoar formation in snow. Thermal processes that are directly affected by these are shown with red dotted arrows labeled as: (*1a*) SOC-induced soil cooling in summer due to the increased insulation to the incoming ground heat flux; *(1b)* SOC-induced soil warming in winter by the reduction of net outgoing ground heat flux; *(2)* heat exchange between shallow and deep soils by shifting the "zero" bottom boundary condition to ~50 m*; (3)* wind speed driven snow compaction, increasing the winter snow thermal conductivity and cooling soils by increasing the outgoing ground heat flux; *(4)* reduction of snow thermal conductivity due to the formation of insulating depth hoar crystals in winter, thereby warming the soils. There are other thermal/hydrological interactions amongst these processes, the effects of which are simulated by the model. The model contains seven above ground vegetation CN pools - four from tree PFTs (foliage, above ground woody biomass, coarse roots, and fine roots) and three from herbaceous PFTs (foliage, coarse roots, and fine roots). These interact with specific atmospheric and biogeophysical drivers and form eight soil/litter CN pools, divided into above ground (metabolic litter, structural litter, microbial soil, humus soil) and below ground (decomposable litter, resistant litter, stabilized humus soil, and microbial soil) components (Yang et al, 2009).

1.1 Evaluating the model estimated terrestrial gross primary productivity using Fluxnet data sets and estimating the climate-driven uncertainties in modeled GPP (Barman et al., 2014a)

We used an Integrated Sciences Assessment Model (ISAM) to quantify the causes and extents of biases in terrestrial gross primary production (GPP) due to the use of meteorological reanalysis datasets. We first calibrated the model using meteorology and eddy covariance data from 25 flux tower sites ranging from the tropics to the northern high-latitudes (Figure 1).

However, substantial uncertainties remain in current model estimates of terrestrial carbon, and it is becoming increasingly necessary to quantify and reduce these uncertainties. One important uncertainty in the models arises from inaccuracies in input datasets itself, such as from meteorological forcings (i.e., climate). While photosynthetic assimilation in land surface models (LSMs) are governed based on mechanistic processes and parameterizations, the climatic/environmental controls determine the specific response and seasonality of ecosystem productivity. For example, climate influences GPP through changes in solar radiation, precipitation, atmospheric temperature and humidity (controls vapor pressure deficit) that determine the supply of light, water and nutrient availability to plant cells. The response of GPP to warming is usually positive at low temperatures and reduces at higher temperatures, and generally increasing with photosynthetically active radiation, and decreasing with increases in vapor pressure deficit. In LSMs with coupled biogeochemistry, the uncertainties in GPP are most likely to produce significant differences in subsequent carbon fluxes (e.g., net primary production (NPP), net ecosystem exchange (NEE), and litter fall), and in soil carbon reservoirs. In addition, given such documented uncertainties in GPP based on meteorological inputs (at regional to global scales), corresponding impacts on modeled derive energy and water fluxes from vegetation also warrant careful study.

We investigate biases in canopy fluxes from reanalysis datasets using ISAM (Figure 1). The current ISAM combines the existing biogeochemical components of the model (Yang et al., 2009; Jain et al., 2009, 2013) with detailed biogeophysical schemes described in Barman et al. (2014a,b), Barman and Jain (2016), El-Masri et al., (2013, 2015), and Song et al. (2013, 2016).

Our specific objectives are as follows: (1) to present the calibration of key vegetation parameters influencing GPP, derived from model optimization at 25 FLUXNET sites; (2) to analyze the climate-driven uncertainties in GPP directly at the site-level using two reanalysis datasets as inputs: CRUNCEP and NCEP/NCAR (references in Methods); (3) to determine the dominant meteorological controls causing the model biases; and (4) to discuss the corresponding impacts on global-scale modeling of GPP. Differing from previous studies based on FLUXNET data using statistical and/or diagnostic techniques, our use of ISAM framework enables us to investigate the causes of these biases due to both environmental (abiotic) and plant functional/physiological (biotic) controls.

Following model optimization, the ISAM estimated annual mean GPP is within the "Mean $\pm$ Uncertainty" estimates from FLUXNET (Figure 2). At most sites, the modeled annual mean GPP difference is $\leq$ 10—15% of the annual FLUXNET mean.

We also compared the mean annual GPP during 2000—2004 from the *ISAM*-NCEP and the *ISAM-CRUNCEP* simulations with two different sources for globally gridded GPP: *(1)* FLUXNET-MTE (Barman et al., 2014a,b), and *(2)* MODIS (MOD17) (Barman et al., 2014a,b) (Figure 3). For consistency, all the GPP estimates (ISAM, FLXNET-MTE and MODIS) were compared based on the $0.5^{o}\times0.5^{o}$ ISAM land mask, and any non-vegetated grid cells were removed from all the sources. Globally, the *relative* GPP difference between *ISAM-NCEP* and *ISAM-CRUNCEP* is only ~4 GtC/yr, the *ISAM-NCEP* being slightly higher than the *ISAM-CRUNCEP*.

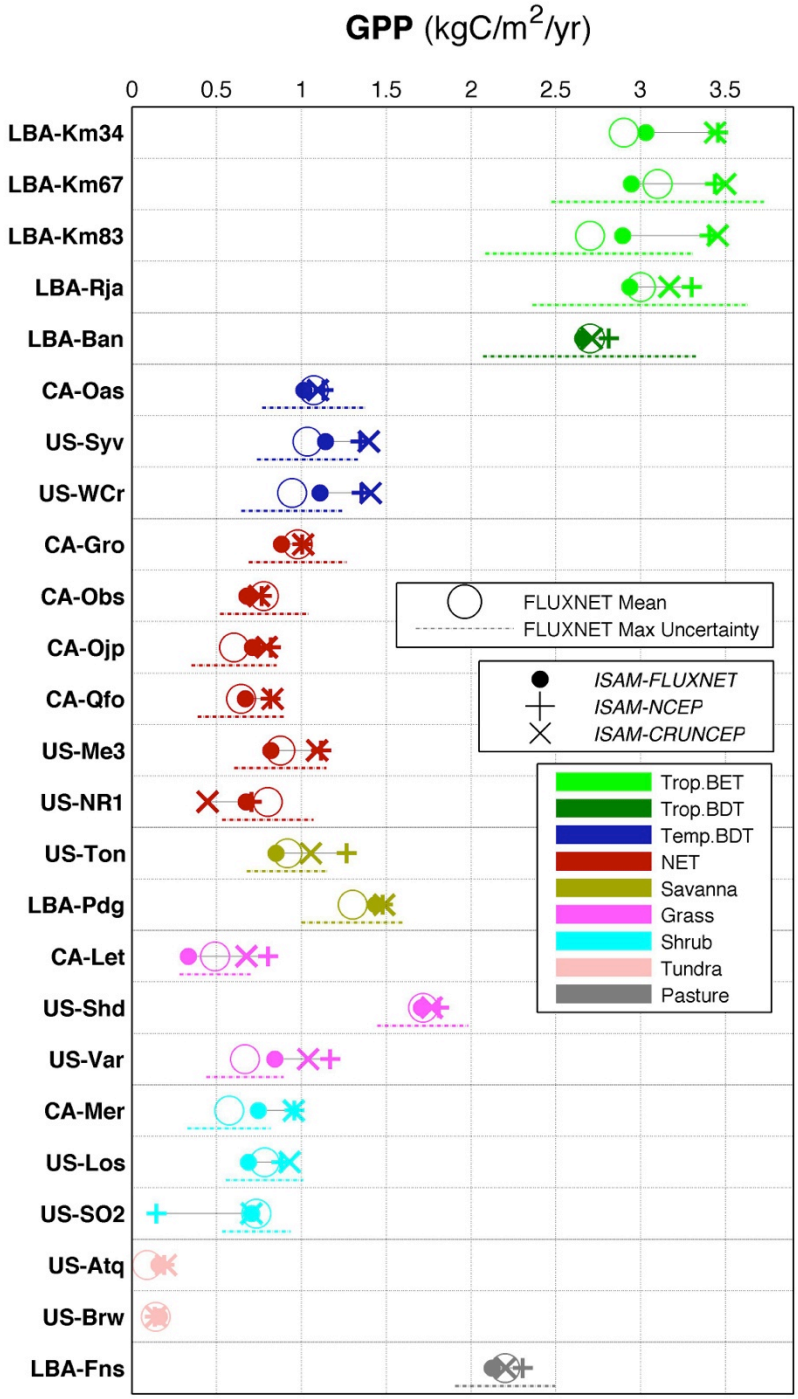


Figure 2: Mean annual GPP at FLUXNET sites used in this study, for observational/flux data and model simulations (*ISAM-FLUXNET, ISAM-CRUNCEP, and ISAM-NCEP*). "FLUXNET max uncertainty" denotes the ±uncertainty range.

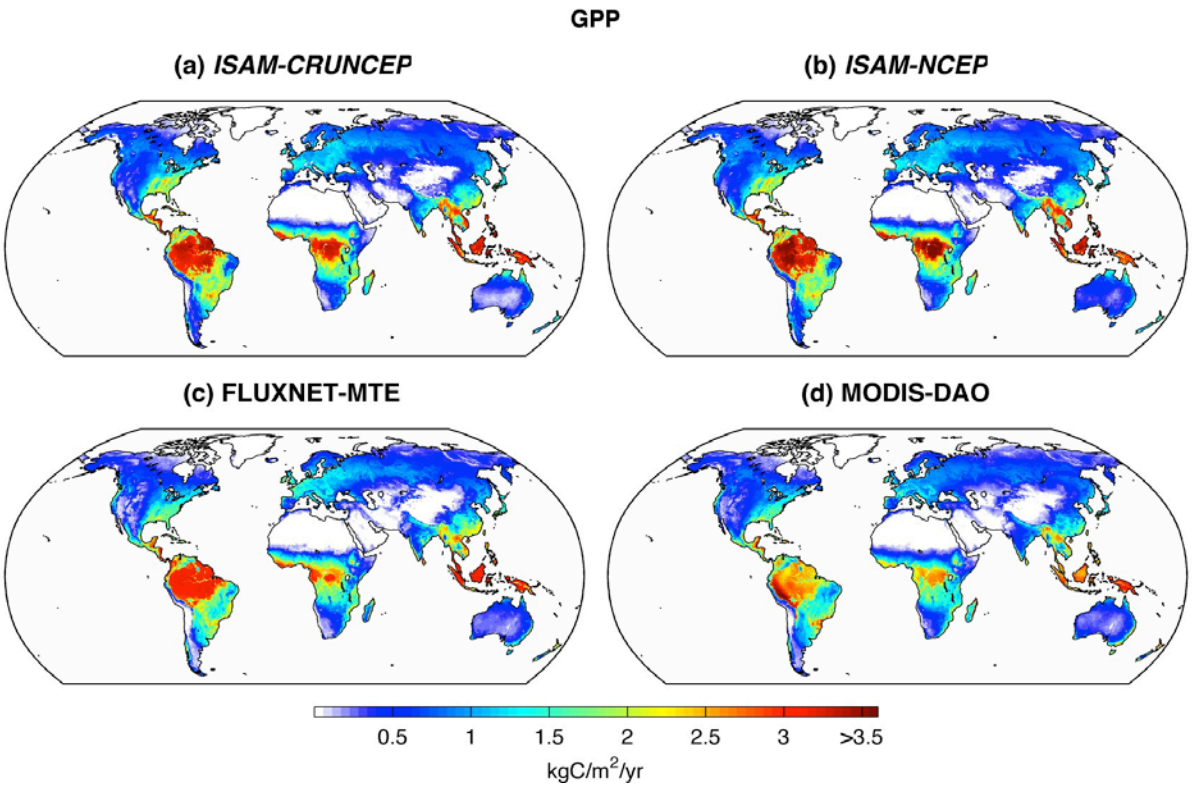


Figure 3: Maps of mean annual estimates of GPP, for two reanalysis-driven model simulations (*ISAM-CRUNCEP* **(a)**, *ISAM-NCEP* **(b)**) and observationally derived datasets (FLUXNET-MTE **(c)**, MODIS-DAO **(d)**). All the results are based on averaged output for 2000—2004, and are only for vegetated land surfaces.

Using the ISAM as the modeling tool, the results show that at most sites (and PFTs) both the NCEP/NCAR and CRUNCEP significantly overestimated the GPP, resulting in ΔGPP of up to ~0.45 kgC/m2/yr for the tropical forest PFTs (Figure 4). For other PFTs, even though the magnitude of the ΔGPP is smaller than those for the tropical forests, the % bias in GPP is significant: i.e., up to +10—20% for the Temp.BDT and NET, and up to ~+20—30% for savanna, grassland and shrubland. Here, one relevant question is, can we choose a preferred reanalysis dataset for model forcing between the NCEP/NCAR and the CRUNCEP? Because both the reanalysis datasets produced similarly (positive) biased GPP, one may not be preferred over the other. Nonetheless, in section 2, we show that the CRUNCEP driven latent and sensible heat fluxes are generally in better agreement than the NCEP/NCAR counterparts when compared with the FLUXNET estimates. Hence, for the overall estimation of carbon, energy and water fluxes, we recommend the use of CRUNCEP data for subsequent global applications of the ISAM.

We recognize several potential limitations in our results, arising due to existing limitations in model structure and parameters. Firstly, due to the use of prescribed LAI climatology in ISAM, it is restricting to simulate the inter-annual variability of GPP. Note that we have addressed this issue in our later study (El-Masri et al., 2015) where we dynamically calculate the plant phenology and LAI. Secondly, the use of static root profiles in ISAM introduces limitations in

modeling of soil water stress, and vegetation acclimation to water stress. Dynamic roots are especially important in the drier tropical and sub-tropical non-tree ecosystems (e.g., C3/C4 grass, C3/C4 savanna, C3/C4 pasture, and shrubs) that have shallow root depths and where roots may vary seasonally. Besides the root profiles, the accuracy of soil moisture schemes used in the model are also important because soil water availability and water stress play significant roles in our results. Currently, there are known limitations of the numerical scheme used in ISAM, which is used in CLM3.5). There are other potentially important climate-driven biotic effects that are not included in this study, such as the temperature acclimation of photosynthesis.

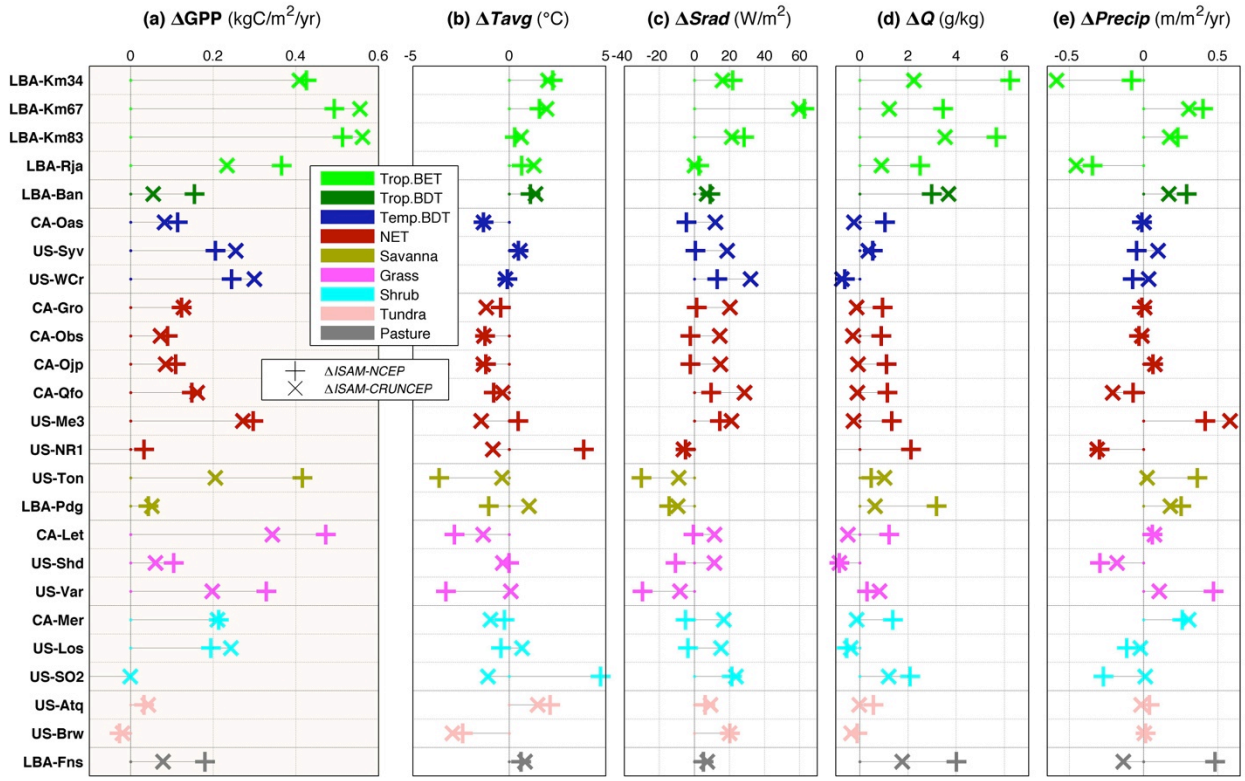


Figure 4: (a) Site-level mean annual GPP biases (ΔGPP) in the reanalysis-driven simulations (*ISAM-CRUNCEP*, *ISAM-NCEP***)**, computed with respect to the respective *ISAM-FLUXNET* simulations. Δ*ISAM-NCEP* = *ISAM-NCEP* – *ISAM-FLUXNET*, Δ*ISAM-CRUNCEP* = *ISAM-CRUNCEP* – *ISAM-FLUXNET*. Note: negative ΔGPP at US-NR1 (using *ISAM-CRUNCEP*) and at US-SO2 (using *ISAM-NCEP*) < -0.1 kgC/m^2/yr were clipped from the Fig. axis. **(b—e)** Mean annual biases in input meteorology variables: **(b)** Δ*Tavg*, **(c)** Δ*Srad*, **(d)** Δ*Q*, and **(e)** Δ*Precip*.

Finally, within the framework of ISAM, the GPP biases are also directly and indirectly coupled to corresponding biases in energy and water fluxes (see next section), with potentially important impacts on soil hydrology and energetics. Uncertainties/biases in these model processes (including their seasonal biases) should also impact other carbon fluxes such as net primary production and net ecosystem production, as well as modeled soil carbon pools. Especially in the mid and high-latitude regions where plant respiration and soil decomposition rates are slower, the extent of GPP biases as shown in this study may be expected to significantly affect the litter fall and the soil carbon accumulation processes. In future studies, it will be important to

investigate these in detail, to explain and to quantitatively reduce the current uncertainties in modeling carbon and energy/water cycle processes in ISAM.

1.2. ISAM estimated terrestrial energy and water fluxes and climate-driven uncertainties in these fluxes (Barman et al., 2014b)

Here, we used a ISAM framework to show that potentially large uncertainties in terrestrial energy/water fluxes can arise from direct biases in reanalysis climate. To consistently quantify the modeled flux biases, we first analyzed the site-level biases in input meteorology from two reanalysis datasets: the NCEP/NCAR and the CRUNCEP. Using these datasets, several consistent patterns in the mean annual ΔR_n, ΔLE and ΔH (Δ: site-level mean annual biases; R_n: net radiation; *LE:* latent heat; *H*: sensible heat) are evident at the site-level (Figure 5) as well as in the global simulations (Figures 6 and 7).

Besides the mean annual biases in meteorology and fluxes, we also highlight the importance of investigating the seasonality of the biases, which have important consequences for the seasonal coupling amongst terrestrial carbon, energy and water fluxes.

With respect to the observed site meteorology, notable biases in the reanalysis variables are as follows: *(1)* high $+\Delta Srad$ (incoming solar radiation) in the tropics (both in NCEP/NCAR and CRUNCEP); *(2)* high $+\Delta LW_{down}$ (incoming downward longwave radiation) in the tropics but negative ΔLW_{down} in the mid/high-latitudes in NCEP/NCAR; and *(3)* high $+\Delta Q$ (specific humidity) in the NCEP/NCAR (Figure 5). There are also varying degrees of biases in *Tavg* (average surface temperature) and *Precip* (total precipitation). Mostly, the mean annual ΔLW_{down} and ΔQ are smaller in the CRUNCEP. Specifically, based on the site-level modeling results of this study, the CRUNCEP driven *LE* and *H* fluxes are generally in better agreement (than the NCEP/NCAR counterparts) with the respective FLUXNET estimates.

We also analyzed the driving factors and mechanisms of the modeled biases in the R_n, *LE* and *H* fluxes. In the model, these flux biases could be primarily attributed to: *(1)* biases in total energy inputs to the surface (*Srad*, LW_{down}); *(2)* biases in *Q* and *Tavg,* which modulate the atmospheric dryness and hence influence *LE/H* partitioning, *(3) Precip,* which may be especially important for non-tree/herbaceous ecosystems. Our model response is typically consistent with several existing observational analysis from literature that suggest, *(1)* R_n controls the seasonal variation of *LE* over the rain forest in Amazonia, and they may not be primarily water stressed; *(2)* available energy is the most important parameter in determining *LE* in the high-latitude boreal forests, which are not predominantly water stressed because of their slow transpiration rates; and *(3)* in arid and semi-arid ecosystems (e.g., non-tree PFTs), *Precip* and factors controlling atmospheric dryness (*Tavg, Q*) are the dominant factors in determining *LE*. However, our results show that the environmental control(s) determining ΔLE can be different from those controlling the response in absolute *LE*. For example, even though R_n may be the dominant factor determining the *LE* response for the tropical forests, the ΔLE is often controlled by ΔQ (unless the positive anomaly in R_n is very high). Here we also note that, as opposed to the usually positive mean annual ΔGPP using the NCEP/NCAR and CRUNCEP datasets as discussed in Barman et al. (2014a), the corresponding ΔR_n, ΔLE, ΔH are not uniformly positive or negative. This suggests that the ΔGPP may not be correlated with ΔLE even though the GPP and *LE* are largely coupled through the stomata. Additionally, because biotic factors such as canopy

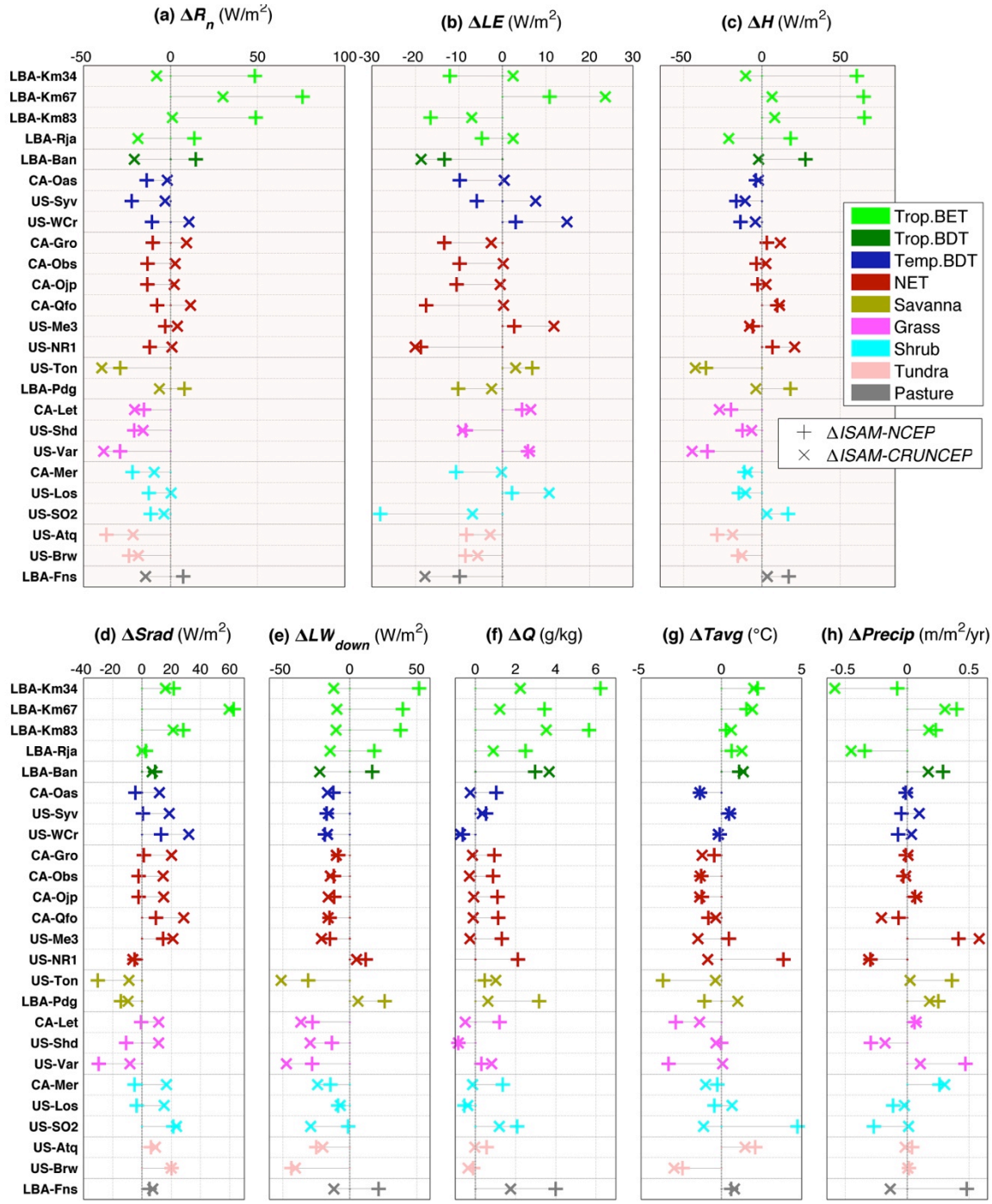


Figure 5: Site-level mean annual biases (Δ) in **(a)** net radiation ($\Delta R_n = \Delta LE + \Delta H$), **(b)** latent heat ($\Delta LE$), and **(c)** sensible heat (ΔH) in the *ISAM-CRUNCEP*, *ISAM-NCEP* simulations. **(d—h)** Mean annual biases in input meteorology variables: **(d)** $\Delta Srad$, **(e)** ΔLW_{down}, **(f)** ΔQ, **(g)** $\Delta Tavg$, and **(h)** $\Delta Precip$. All the biases were calculated with respect to the *ISAM-FLUXNET* counterpart.

physiology/morphology, and environmental factors such as soil thermal and hydrological processes also determine the *LE* and *H* response (and biases) in the model, any *single* factor individually should *not* be expected to fully explain the energy/water flux biases and the variations in model response.

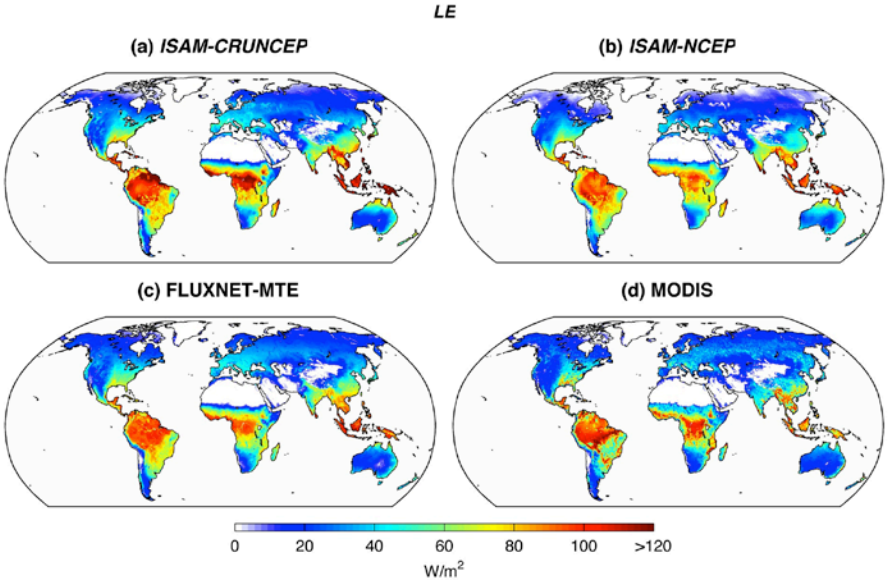


Figure 6: Maps of mean annual estimates of *LE*, for two reanalysis-driven model simulations (*ISAM-CRUNCEP* **(a)**, *ISAM-NCEP* **(b)**) and two observationally derived datasets (FLUXNET-MTE **(c)**, MODIS **(d)**). All the results are based on averaged output for 2000—2004, and are only for vegetated land surfaces.

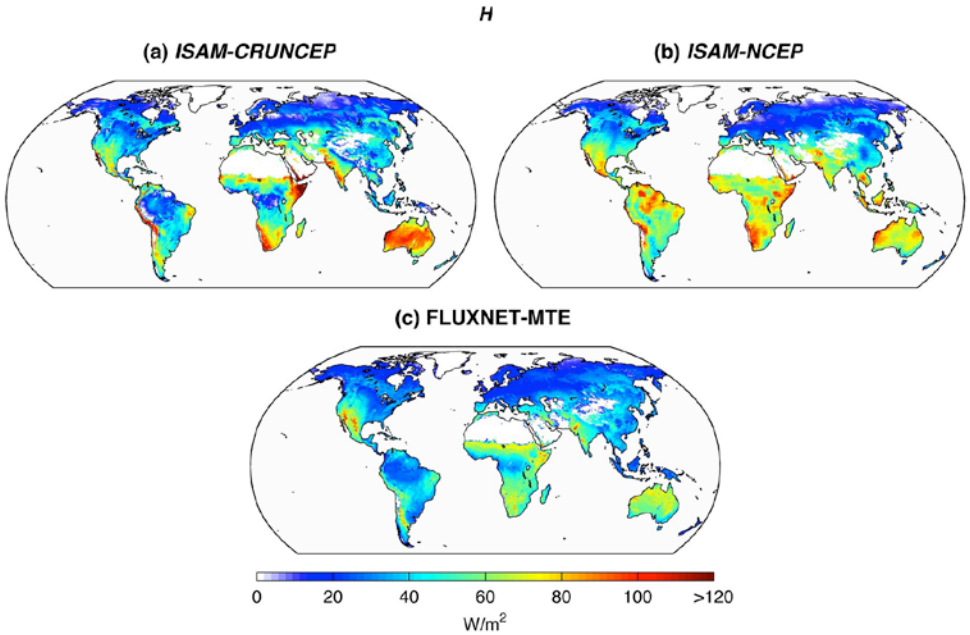


Figure 7: Maps of mean annual estimates of *H*, for two reanalysis-driven model simulations (*ISAM-CRUNCEP* **(a)**, *ISAM-NCEP* **(b)**) and an observationally derived dataset (FLUXNET-MTE **(c)**). All the results are based on averaged output for 2000—2004, and are only for vegetated land surfaces.

LSMs are ultimately designed to study complex land-atmosphere interaction processes, and for application into future climate/environmental change scenarios, at regional to global scales. Hence, given the magnitudes of biases in the *LE* and *H* fluxes using reanalysis data, we feel that further study is warranted to quantify the associated impacts on various land-atmosphere exchange parameterizations dependent on partitioning of energy fluxes. Also, several important questions do arise on the philosophy of LSM calibration using site-scale FLUXNET data. For example, what are the net impacts of site-level calibration on *global* estimates of various fluxes and reservoirs from LSMs, and should we calibrate to optimize at the sites or should we rather optimize the model based on global datasets to counteract such biases in the first place? In this context, we argue that the only reliable alternative may be to improve the global reanalysis products to consistently force the model simulations. As for the current reanalysis products, it may be useful to develop potential strategies to indirectly account for the LSM output biases post model simulations – such as the scaling of fluxes to account for the established model biases. For such purposes, estimation of site-level flux biases as demonstrated in our study (Barman et al., 2014a) is a useful first step to formulate the respective scaling factors.

While continued model evaluation is necessary to improve the representation of carbon, energy and water cycles in the ISAM, here we demonstrate the need to systematically investigate the flux uncertainties from forcing datasets itself, such as from meteorology. Better quantification of uncertainties should lead to better attribution of uncertainty sources, which can ultimately help to reduce the errors in future modeling efforts. Because LSMs usually use many similar schemes across models (due to shared model development or through infusion of sophisticated schemes from other models when available), the magnitude and range of flux uncertainties presented in this study is expected to be of interest to other LSM modelers, and to the ESM community in general.

1.3. Implementation of a Dynamic Rooting Depth and Phenology into a Land Surface Model: Evaluation of Carbon, Water, and Energy Fluxes in High Latitude Ecosystems (El Masri et al., 2015)

The northern high latitude ecosystems (NHLEs) have experienced rapid warming over the recent decades, in response to increased air temperature, and the same trend is expected throughout this century. Some of the major responses of the NHLE to warming include the changes in carbon (C) storage in vegetation and soil, mainly as a result of advancement in the onset of the NHLE growing season, shifts in ecosystem composition and abundance, such as higher shrub density, and latitudinal advances in the tree line. Given these changes, it is important to understand and quantify the response of the NHLE to changes in C, water, and energy fluxes under climate change.

While continued research on the development of more detailed Earth system models (ESMs) is essential to understand the interactions and feedbacks among vegetation, soil, and climate change in the NHLE, one of the challenges for the Land Surface Models (LSMs) as we discussed in previous two sections and our papers (Barman et al., 2014 a,b), a land component of the ESMs, is the treatment of the exchanges of C, water, and energy between the terrestrial ecosystems and the atmosphere and the impact of these exchanges on climate through a variety of biophysical and biogeochemical pathways over a range of spatial and temporal scales.

The structural properties of vegetation, particularly the phenology and rooting depth, are important to study the exchanges of C, water, and energy between the terrestrial biosphere environment of the NHLE and the atmosphere. Additionally, changes in structural properties can also impact the global and regional climate due to their feedbacks with environmental factors, such as atmospheric CO2 concentration. Modeling the vegetation structural properties is more challenging in the NHLs, because of the scarcity of observational data to evaluate the model performance. While most of the previous modeling studies accounting for deep roots have focused on the tropical evergreen forests, there is also a need to study the impact of deep roots and dynamic roots on GPP and water uptake in the NHLs. It is also important to note that in many LSMs the allocation of assimilated vegetation C to leaves, stems, and roots is based on fixed fraction of assimilated C and can limit the models ability to predict the vegetation response to changes in key environmental variables, such as soil nitrogen.

Most LSMs also do not properly include a strong coupling between C-N processes with soil/snow hydrology, and energy [McGuire et al., 2006], which is another factor that impacts the vegetation structural properties and SOC in the NHL regions. Thus, the objective of this study is to implement the dynamic vegetation structure properties in ISAM to understand how these properties are changing due to changes in environmental factors that impact the C, water, and energy fluxes. This study builds upon and extends the application of the dynamic phenology and the rooting depth and distribution in ISAM, by implementing a new dynamic phenology component and dynamic root parameterizations in the NHLs.

The new implemented parameterizations differ in many ways from the other LSMs. For instance, the new parameterization account for the impact of the soil moisture stress on leaf onset, which is not only important in NHLEs, but also is important in subtropical herbaceous ecosystems, because soil moisture demand triggers the growing season. In addition, the timing of the leaf onset and offset are determined by the temperature and day length thresholds, which are particularly important in the temperate and high latitude ecosystems where shorter and colder days induce dormancy. Unlike many other LSMs, the dynamic phenology in ISAM takes into consideration the nutrient availability, particularly N, on the dynamic N allocation to leaf, stem and roots. Furthermore, the linkage between phenology and plant water uptake are emphasized by highlighting the role of dynamic rooting depth and distribution as an essential scheme in LSMs for a better prediction of the growing season length in the NHLEs. The dynamic rooting depth and distribution allow the plant to grow its roots laterally and horizontally to maximize the access to soil moisture, altering the water stress on plant productivity and improving the model LAI and growing season length estimation.

Using these new schemes, ISAM performance was evaluated through the interactions between vegetation growth and C, water, and energy fluxes with observational data in the NHLEs. Errors in the modeled fluxes were also estimated with and without including dynamic processes. Here we describe the major findings of our modeling analysis, the detailed study findings can be found in El Masri et al. (2015).

The dynamic phenology and dynamic rooting depth and distribution were implemented and calibrated into ISAM in four NHL ecosystems (NHLEs) that are representative of the major biomes in the NHLs poleward of 45^{o}N. Also, the new dynamic version of ISAM model,

$ISAM_{DYN}$, was validated at multiple boreal evergreen, one boreal deciduous and one tundra sites. The cumulative root fractions in the NHLEs calculated based on the $ISAM_{DYN}$ captured well the trend in the measured root fraction with soil depth, while the root fraction on the original or base case, $ISAM_{BC}$, showed only a fixed curve shape (Figure 8). The $ISAM_{DYN}$ root depth was in close agreement with measurements (Figure 8), because the dynamics parameterization more realistically represented the vertical root growth that was initiated in response to water or nutrient stresses, compared to the fixed root fraction in the $ISAM_{BC}$.

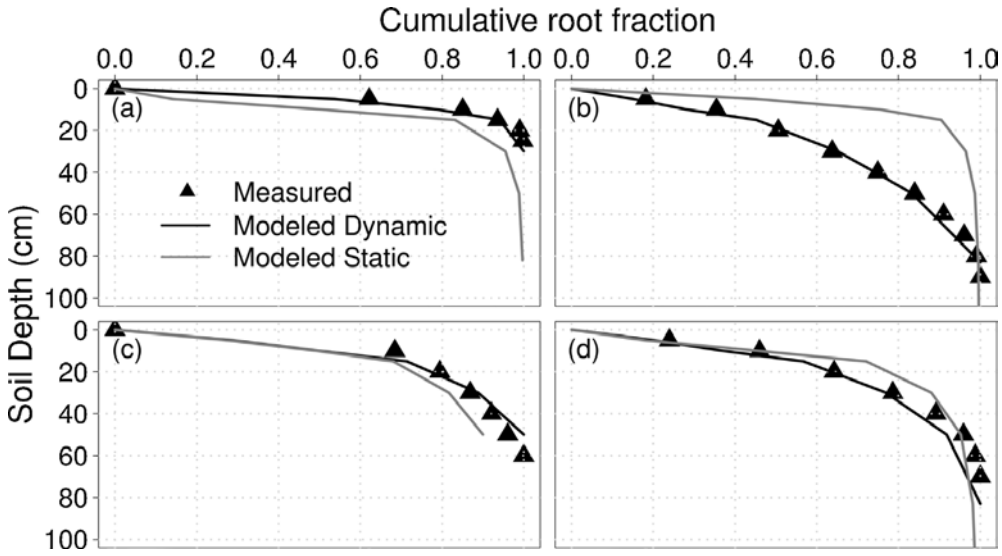


Figure 8: Comparison of dynamic ($ISAM_{DYN}$) and static ($ISAM_{BC}$) models estimated cumulative root fractions for four calibration sites: (a) tundra (Barrow, AL; lat=71.32N, lon= 156.82W); (b) grassland (Dickinson, ND; lat=46.9N, lon = 102.82W); (c) boreal deciduous (Alberta Aspen, Canada; lat= 55.28N, lon= 114.77W); and (d) boreal evergreen (Alberta Black spruce, Canada; lat=55.28N, lon = 114.77W).

The $ISAM_{DYN}$ LAIs, particularly the timing and the magnitude of the peak LAIs matched the measured seasonal variations in the NHLE, whereas the MODIS LAI data failed to capture leaf onset and senescence, and the annual variation of LAIs as compared to ground based measurement (Figure 9).

Model results suggests that as a result of the improvements of modeled phenology and root density and depth, the $ISAM_{DYN}$ is able to capture the seasonal variability in the GPP, LH, and SH fluxes through both direct and statistical comparisons to site measurements (Figure 10 and 11). Overall, the $ISAM_{DYN}$ GPP and LH were in better agreement with site flux tower data compared to the $ISAM_{DYN}$ SH. Our site data analysis suggested that these sites have energy balance closure errors. This might have been one of the reasons that the model results in some cases might not have been consistent with the flux tower data. Further analysis into the model biases in the partitioning of energy fluxes is needed to assess the biases in the modeled SH.

The relevance of this experiment is the linkage between the dynamic vegetation structure processes (dynamic phenology and dynamic rooting depth and distribution) and how that linkage improves LAI simulations. Moreover, the results demonstrated how vegetation structure

processes alongside the dynamic C allocation impacted the model simulated C and energy fluxes. Model experiments performed in this study indicated that excluding any of these dynamic processes resulted in higher biases in the C and energy fluxes for most of the study sites. Therefore, the importance of including these dynamic processes in LSMs is emphasized, and more importantly for the LSMs that calculate LAI, not only to improve the modeled C and energy fluxes, but also to capture the observed vegetation seasonality and GSL. Because most LSMs are capable to estimate annual GPP that is consistent with flux tower data, but fail to simulate the observed GPP seasonality. This study is a step toward advancing model predicted seasonality.

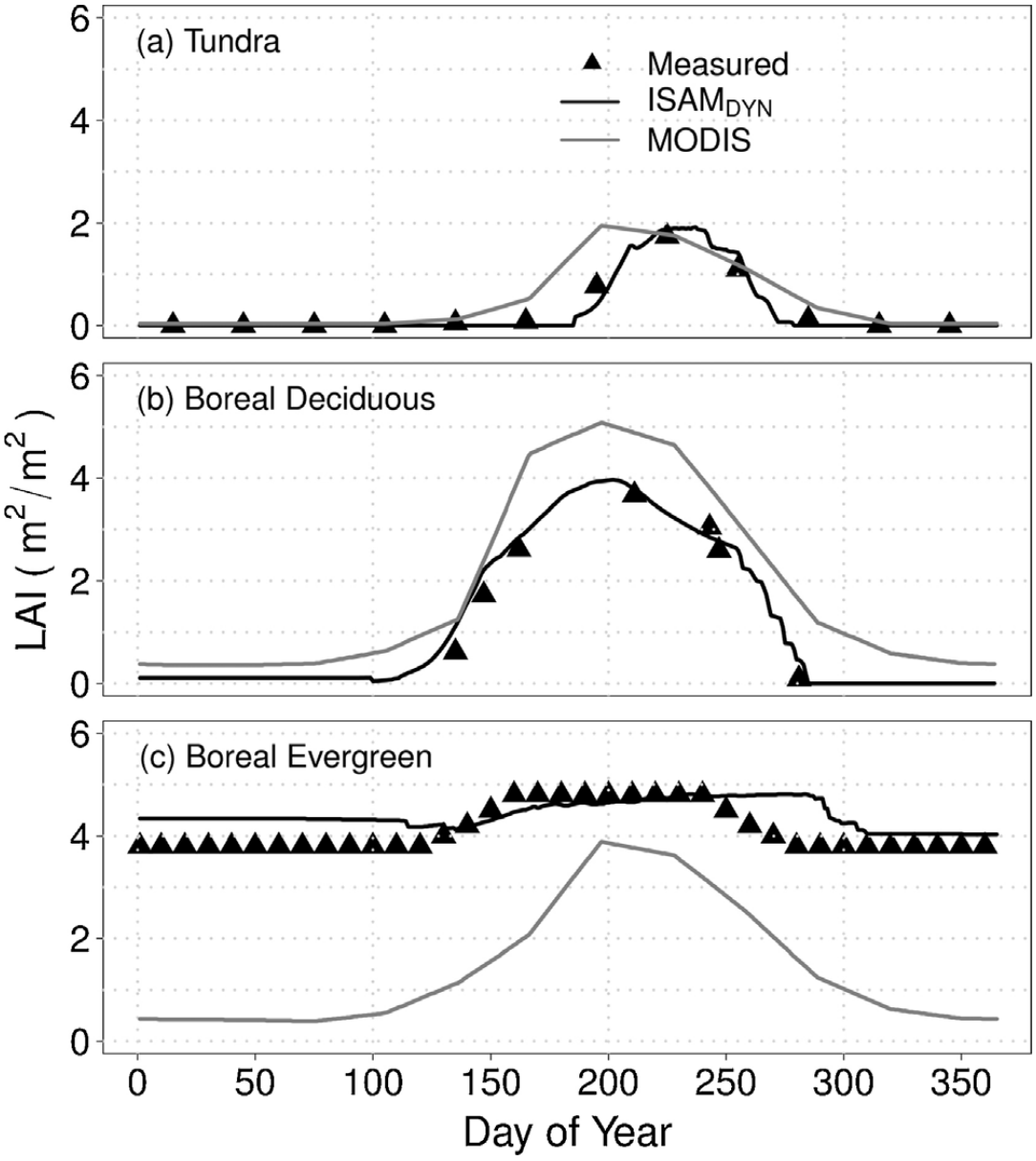


Figure 9: Comparison of dynamic (ISAM$_{DYN}$) model and MODIS estimated LAI for (a) tundra (Barrow, AL; lat=71.32N, lon= 156.82W); (b) boreal deciduous (Oas; lat= 55.63N, lon= 106.19W); and (c) boreal evergreen (Nobs; lat=55.88N, lon = 98.48W.).

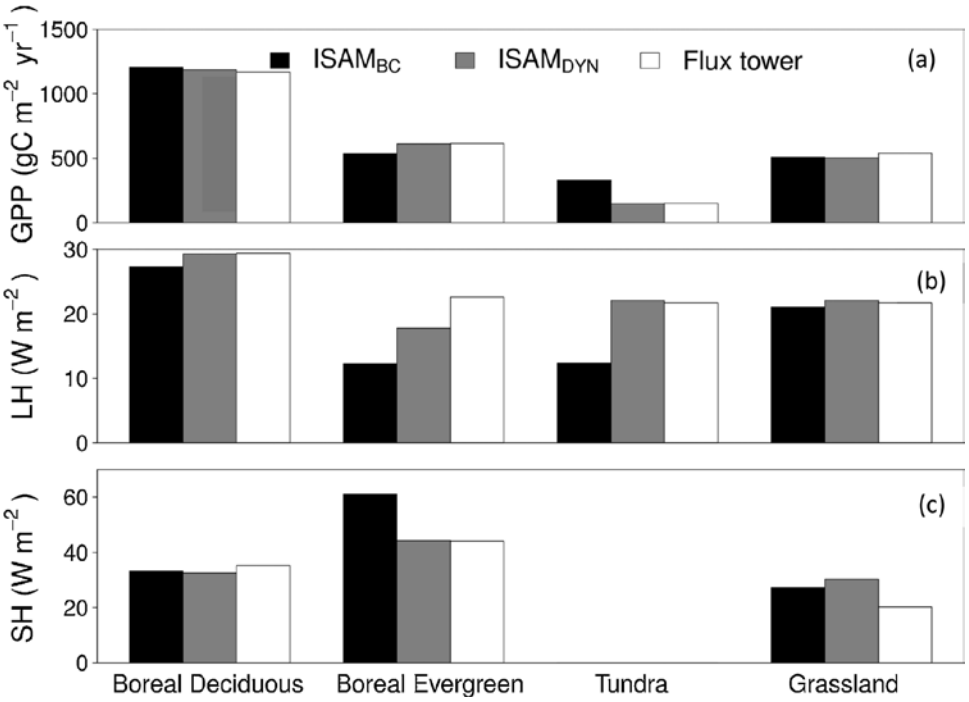


Figure 10. Comparison of dynamic (ISAM$_{DYN}$) and static (ISAM$_{BC}$) models estimated annual (a) GPP, (b) LH, and (c) SH (ISAM$_{DYN}$, ISAM$_{BC}$) with flux tower measurements for the calibrated sites in the northern high latitude region (Oas, Nobs, Barrow, Let). The two version of the model results were not compared with flux tower SH data for Tundra biome, as no data were available.

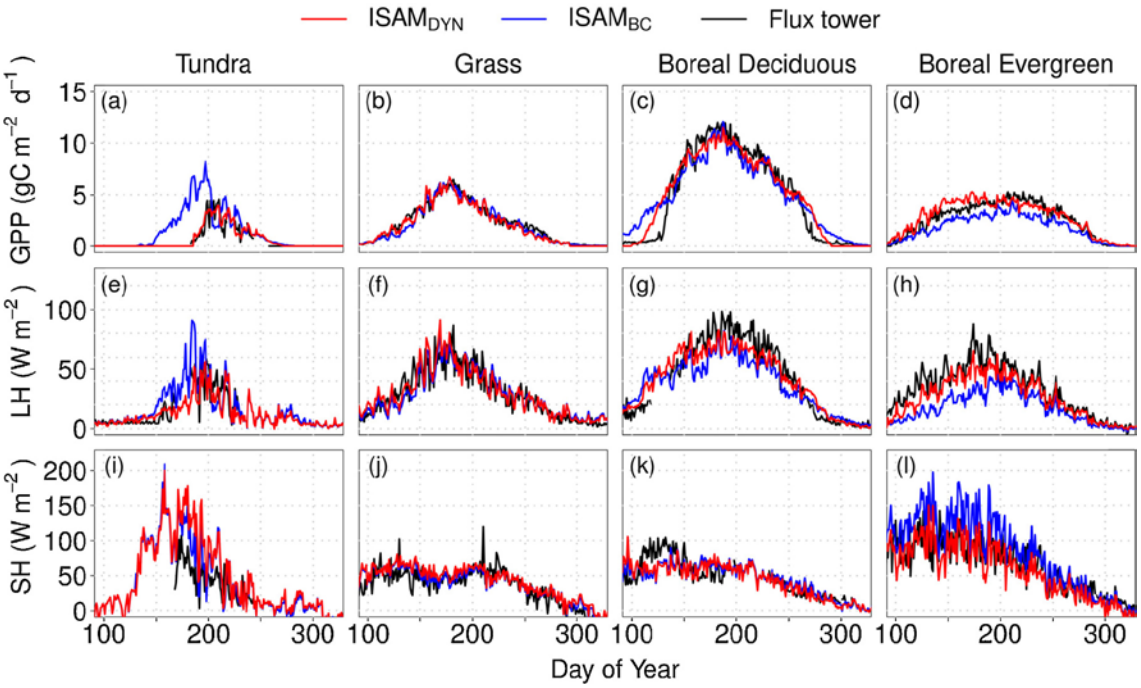


Figure 11: Seasonal variability of daily tower fluxes and model simulated GPP, latent energy (LH), and sensible energy (SH) for the calibrated sites.

The implementation of the dynamic phenology and rooting distribution allowed the $ISAM_{DYN}$ to better quantify the feedbacks between climate change and C and energy fluxes. These processes helped to simulate the vegetation adaptation to changing climate, because of adjustments in the plant phenology and structure. For instance, the implementation of these dynamic processes resulted in significant improvement in the tundra simulated C and energy fluxes. Thus, the lack of these processes in LSMs may lead to the NHL C and energy fluxes overestimation, causing errors in LSMs response to climate change.

1.4 Effects of cold-region soil/snow processes and the uncertainties from model forcing data on permafrost physical characteristics (Barman and Jain, 2016)

With the NHL regions warming continually, it is increasingly important to quantify the thermal state of current permafrost as well as its future degradation. While there is a consensus that near-surface permafrost area will continue to decrease with climate warming, the rate of degradation produced by modeling studies remain highly divergent. Changes in permafrost can impact regional terrestrial energetics, hydrology, and ecology; consequently, large-scale permafrost thaw is expected to mobilize the soil carbon and tremendously impact global climate. However, in marked contrast to their recognized importance, our understanding of permafrost and observations of high-latitude soil/snow processes remain sparse.

An underlying difficulty arises in modeling the contemporary Northern Hemisphere near-surface permafrost area itself, as evident from the diagnosed range of 1.5—27.3 million-km^2 (during 2005) across the recent Coupled Model Intercomparison Project (CMIP5) earth system models (ESMs). Subsequent model diagnoses have attributed several deficiencies in model structure and parameterizations, mostly related to representation of soil and snow thermal processes such as: (1) thermal coupling between deep and shallow soils, (2) impact of soil properties due to organic content, (3) representation of snow physics and insulation processes, (4) inclusion/exclusion of energy transfer from phase change, (5) interactions between soil energetics and hydrology, etc. In a number of studies, the authors noted that the majority of current models do not represent many of these processes, causing a wide range in simulated permafrost area and degradation rate. In addition, the climate forcing (meteorology) can also influence the permafrost, through the modification of energy/water exchange between the atmosphere and the soil surface. Across currently available ESMs, the computed meteorology can be highly divergent. Similarly, in simulations using land surface models (LSMs) the meteorology from one of many available reanalysis can also be sufficiently different, which can cause divergent permafrost estimates. Some studies investigated the role of climate biases on the diagnosed permafrost in CMIP5 ESMs, by using diagnostic indices to isolate the contributions of model-simulated climate on permafrost, from those due to model structure. Their analysis shows that biased climate can significantly degrade permafrost predictions. However, given their use of indirect and simplified indices for this analysis, a direct estimation of permafrost sensitivity to meteorology/climate (such as in a LSM framework) remains necessary. Furthermore, the modeling of permafrost physical characteristics in the NHL are most likely to suffer from considerable differences in current land-cover datasets. The extent of impacts from such uncertainties in tandem with model (soil/snow) structural differences has not been quantified in existing literature.

Based on the lessons learned from previous studies, in this study we performed an integrated analysis using ISAM, to (1) study the sensitivity of permafrost physical characteristics to specific

improvements in cold-region soil/snow thermal processes, and (2) compare the importance of such changes in model structure and parameterizations with modeling uncertainties from climate and land-cover datasets. Specifically, model structural improvements are represented using two soil and two snow processes that are prevalent in the NHL environment: (1) energy exchange between shallow and deep soils, by representing soils up to ~50 meters, (2) effect of soil organic carbon (SOC) on thermal/hydrological properties, (3) wind compaction of snow depth (and density), and (4) depth hoar formation in snow. A conceptual diagram of these processes and how they interact with permafrost physical characteristics are shown in Figure 1. Extending from previous studies, a key goal of this specific study was to quantify the individual contribution of these soil/snow processes on multiple variables: permafrost extent, area, degradation, soil thermal and hydrological states for the entire NHL. The responses from these soil/snow improvements are then contrasted with driver-induced uncertainties from two reanalysis (climate) datasets and two land-cover datasets cover within the single unified model domain. Finally, we also investigated the sensitivity of permafrost area to a seemingly standard factor in model diagnosis of permafrost – the choice of threshold soil temperature (by default chosen as mean monthly temperature of 0^{o}C), the results of which yield interesting insights.

In the following, we discuss the implications of our results in the context of divergent permafrost area estimates from multi-model intercomparison projects, and soil biogeochemical calculations.

We show that cold-region soil/snow biogeophysics can strongly affect the simulation of NHL permafrost area, degradation, and soil thermal and hydrological states in land surface models (Figures 12 and 13).

Table 1. List of Model Experiments and Corresponding Permafrost Area (PA)[a]

	MODEL			CLIMATE		LC		
Experiment	New	Old	Interm	CRUNCEP	NCEP	LC1	LC2	PA 2000–2004 ($\times 10^6$ km^2)
NEW	✓			✓		✓		12.3
OLD		✓		✓		✓		2.9
NEW-NO-DH			✓	✓		✓		13.0
NEW-NO-DS			✓	✓		✓		8.7
NEW-NO-SOC			✓	✓		✓		9.6
NEW-NO-WIND			✓	✓		✓		10.3
$NEW_{CRUNCEP-LC2}$	✓			✓			✓	12.1
$OLD_{CRUNCEP-LC2}$		✓		✓			✓	3.1
$NEW_{NCEP-LC1}$	✓				✓	✓		10.5
$OLD_{NCEP-LC1}$		✓			✓	✓		2.6
$NEW_{NCEP-LC2}$	✓				✓		✓	9.9
$OLD_{NCEP-LC2}$		✓			✓		✓	2.7

[a]Soil/snow physics (*MODEL*) is classified as *New*, *Old*, and *Interm* (intermediate). Each *Interm* version was created by excluding one process from *NEW* at a time, as indicated by suffix: (1) *NO-DH*: no depth hoar, (2) *NO-DS*: no deep soil, (3) *NO-SOC*: no SOC, and (4) *NO-WIND*: no wind compaction of snow. *Old* was created by excluding all the aforementioned processes from *New*. For different simulations, the choice of meteorology (*CLIMATE*) was between the CRUNCEP and the NCEP, and the choice land-cover data set (*LC*) was between the LC1 and the LC2. Unless otherwise specified through the subscript, a simulation (i.e., the first six simulations) was driven using the CRUNCEP and the LC1.

With respect to observations, there can be strong negative biases in the simulated permafrost area when critical soil/snow processes are excluded in the model. Much of these biases can be corrected by their inclusion. Many of these processes/parameterizations are currently missing even in the current generation of LSMs and ESMs. Therefore, as demonstrated here such models are likely to benefit by including them in their land surface schemes. Besides the modeling

improvements, we discuss limitations in the model, such as the lack of sub-grid variability in representing permafrost classes with less than 50% grid area coverage, limited land-cover classification types, and other factors. Necessary parameterizations for many such processes are still being developed in the community. When they are available in the future, they can be expected to further improve the permafrost physical characteristics in the model.

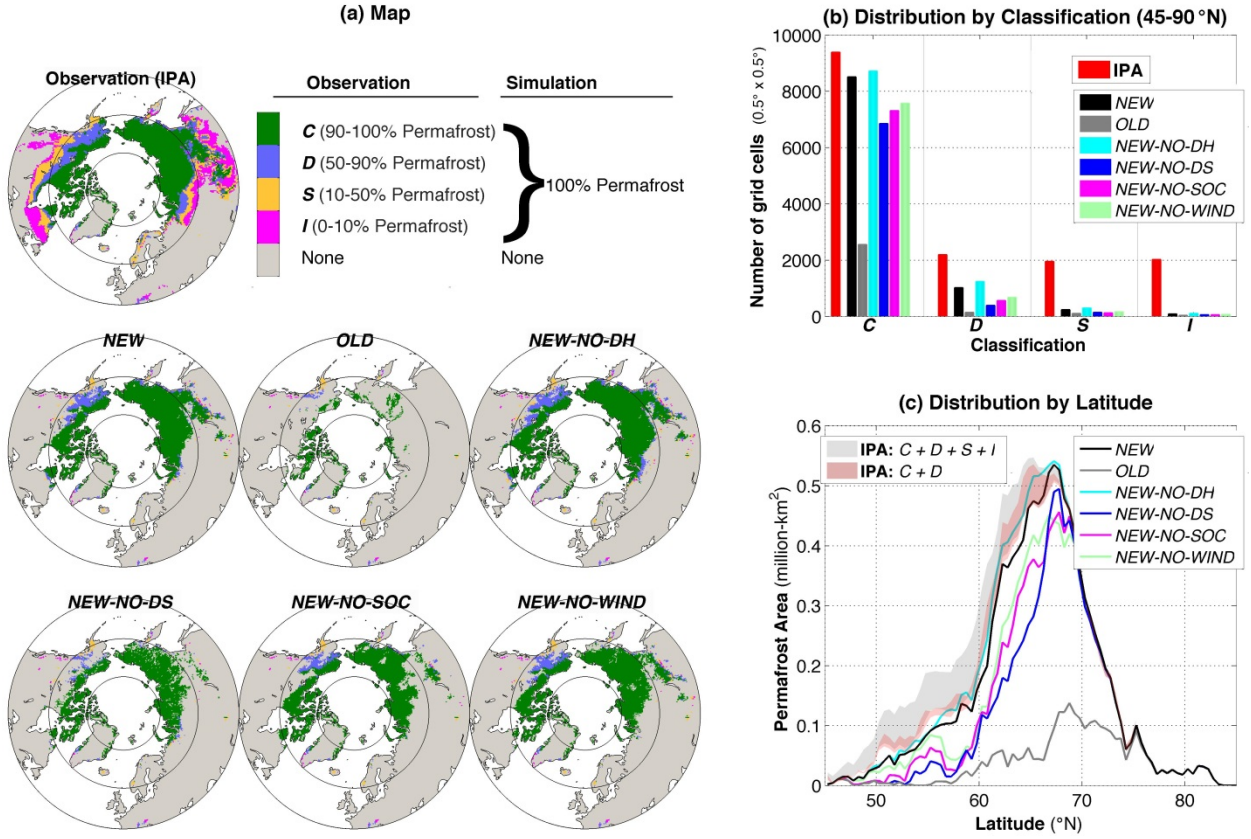


Figure 12: (a) Map of permafrost area (poleward of 45°N) by latitude in observations (International Permafrost Association: IPA), and simulations: *NEW*, *OLD*, and four *Interm* versions (*NEW-NO-DH*, *NEW-NO-DS*, *NEW-NO-SOC* and *NEW-NO-WIND*) (See Table 1 for various simulation cases). In observations, different permafrost classifications based on % grid cell coverage are: Continuous (*C*): 90—100%, Discontinuous (*D*): 50—90%, Sporadic (*S*): 10—50% and Isolated (*I*): 0—10%. In simulations, a grid cell is either 100% permafrost or no permafrost. For gridcells that are diagnosed as permafrost (i.e., 100% coverage) in respective simulations, the colors illustrate the observed pemafrost type therein. **(b)** Number of $0.5^{o} \times 0.5^{o}$ grid cells within each classification in observations, and subset of respective grid cells diagnosed as permafrost in simulations. **(c)** Distribution of permafrost area by latitude. The range of observated permafrost area shown in shading was derived by weighting fractional coverage of different permafrost classifications. All the simulated results are using averaged output during 2000—2004.

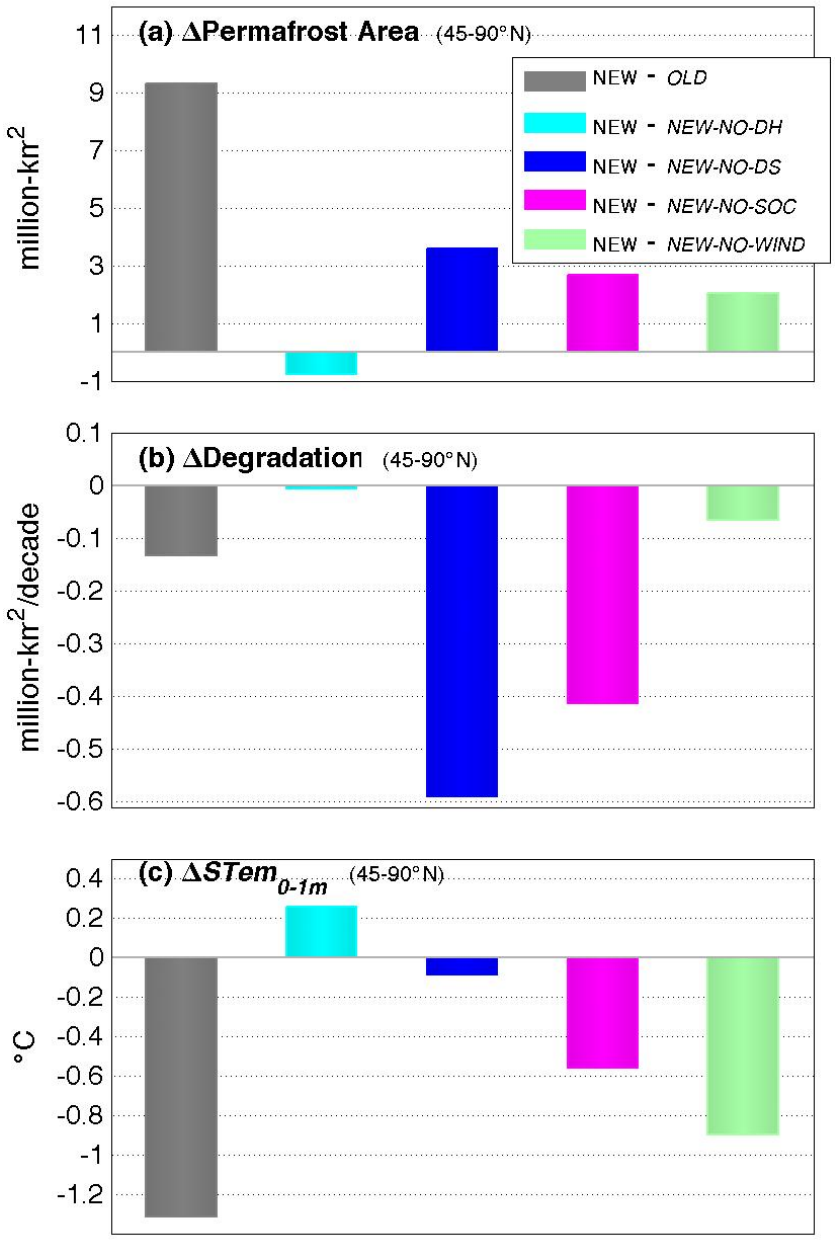


Figure 13: Simulated differences (Δ) in key permafrost physical characteristics in *OLD* and four *Interm* simulations with respect to *NEW*: in mean annual **(a)** Permafrost Area (averaged during 2000—2004), **(b)** Degradation rate (from 1985—2010), and **(c)** Soil temperature in 0—1 m depth ($STem_{0\text{-}1m}$, averaged during 2000—2004).

A specific goal of this study was to quantify the importance of each of the four soil/snow processes on permafrost physical characteristics, including soil thermal and hydrological states (Figure 14).

In diagnosing the permafrost area and degradation (for 0—3 m soil column), the strongest contribution of deep soil thermal dynamics is evident, followed by insulating properties from

SOC, wind compaction of snow, and depth hoar formation. However, the importance of these processes on soil energetics and hydrology changes with depth. Most importantly, in the soil column dominated by vegetation root zone (~0—1 m), the influence of deep soils becomes minimal. Biogeochemical activities in soils, which are primarily active within the root zone and dependent on soil temperature/moisture therein, are therefore likely to be minimally influenced by inclusion of deep soil thermal dynamics. In this context we show that wind compaction of snow, SOC-induced modifications in soil properties and depth hoar formation – all of which are prevalent in the high-latitude environments, influence the root zone temperature/moisture to a great extent. Therefore these can be expected to improve model estimation of permafrost SOC within the top 1 m of soil column. Note that as also shown in this study, near-surface permafrost area in the model can be highly sensitive to the threshold for soil temperature to diagnose permafrost. But, because the choice of this threshold doesn't affect the simulated soil states itself, the diagnosed NHL permafrost area in a model may be decoupled from the corresponding estimates of NHL SOC (i.e., more permafrost area may not imply more simulated SOC in the NHL).

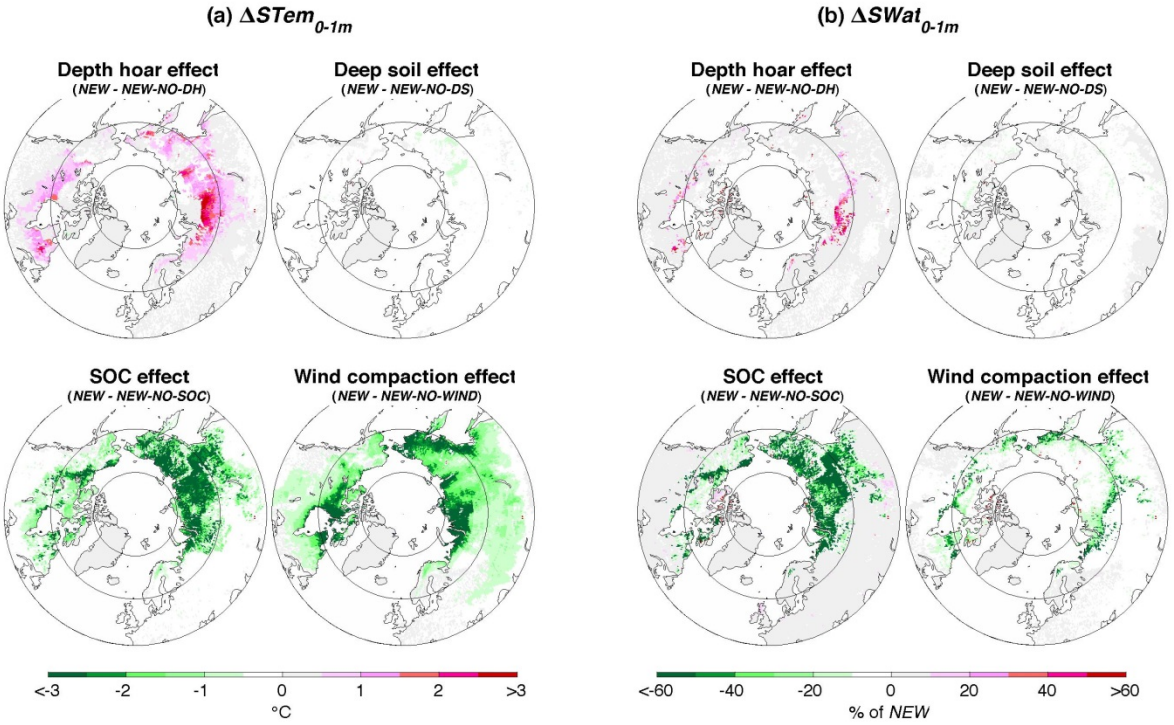


Figure 14: (a) Simulated effects on annual soil temperature in 0—1 m depth ($STem_{0\text{-}1m}$) due to individual *inclusion* of: depth hoar formation, deep soils, effects of SOC on soil thermal/hydrological properties, and wind compaction of snow. **(b)** Corresponding effects on annual soil water in 0—1 m depth ($SWat_{0\text{-}1m}$). The results shown are for averaged period during 2000—2004.

Besides the contribution from soil/snow processes, we also show that climatic uncertainties in datasets can produce notable differences in NHL permafrost area (e.g., a difference of ~0.5 ^{o}C in mean annual air temperatures between the two reanalyses lead to a permafrost area difference of 2.3 million-km^2). Given that much larger meteorological differences occur across currently

available reanalysis datasets, we argue that climate-driven uncertainties are likely to play a much greater role when using these datasets. The implications of even greater differences in surface meteorology across models in multi-model intercomparison projects are therefore likely to be tremendous for permafrost area and thermal state, both for the present and for the future. This should also be true for NHL energy/water fluxes, which are influenced the least from soil/snow changes investigated here, but are majorly impacted by differences in climate and/or land-cover.

1.5 Dominant role of soil physical states over nitrogen cycle on permafrost soil carbon

The permafrost carbon feedback in the northern high-latitudes (NHLs) will be governed by the amount of soil organic carbon (SOC) currently stored therein, and how it responds to the changing climate. The extent of this potentially strong feedback can only be calculated using tools, such as earth system models (ESMs). However, permafrost SOC estimates from current generation of ESMs are strongly divergent, such as 60—820 PgC in top 1 m of soil across the CMIP5 models. In another multi-model intercomparison project – the MsTMIP, global SOC distribution across terrestrial biosphere models (TBMs) also appear to be most divergent in the 40—65^{o}N latitudes.

The difficulty in modeling SOC in permafrost regions stems from two challenging aspects: *(1)* the complexity of carbon-nitrogen (CN) biogeochemistry in the high-latitude ecosystems, and *(2)* a large suite of active biogeophysical processes in areas affected by the permafrost. In this context, while continued development of detailed CN biogeochemistry is necessary for the NHL environment, a key challenge is to model cold-region soil/snow physics that critically influence the soil hydrology and thermal energy, such as shown in previous studies: deep soil thermal dynamics, feedback of SOC on soil thermal/hydrological properties, snow insulation effects from depth hoar formation and wind compaction of snow, etc. Most current TBMs have not yet incorporated them. Previously, divergent SOC estimates across TBMs have been attributed to differences in representation of temperature/moisture sensitivity, nutrient limitation, land-cover, etc.. Given the coupling between soil hydrology, energy and CN processes (e.g., N-limitation or fertilization on vegetation growth, litter and soil decomposition), cold-region soil/snow physics is most likely to play an important role in the calculation of permafrost SOC. Lastly, uncertainties in climate and land-cover datasets introduce uncertainties in the calculations of NHL SOC. However, the relative impacts of these factors within a TBM remain unknown, and we argue that quantifying these is an important step towards simulating more accurate permafrost SOC, and the associated carbon-climate feedback.

Here using a TBM with CN biogeochemistry, we performed such an analysis to address the following questions. *(1)* What are the relative strengths of cold-region soil/snow physical processes, N-cycle, and uncertainties from climatic and land-cover drivers on permafrost SOC buildup in the NHLs? *(2)* What are the impacts of individual mechanisms (e.g., changes in soil temperature, moisture, litter input, etc.) on SOC, and which dominate and where? *(3)* Finally, what are broad significances of our results in the context of explaining divergent SOC estimates from contemporary models, and associated implications for improving current model/data products? We focus on SOC within 0—1m of soils (subsequently, $SOC_{0\text{-}1m}$), and represent cold-region with four soil/snow processes specified in the previous paragraph: deep soils, SOC-induced modification of soil physics, wind compaction of snow, and formation of depth hoar crystals – all of which impact soil states and CN dynamics (Figure 1). The impacts of CN

dynamics are tested using N-limitation vs. no N-limitation scenarios, and climate and land-cover data induced uncertainties are tested using two datasets of each.

We begin with results for soil/snow physics driven $SOC_{0\text{-}1m}$ changes from two model experiments: *NEW* – the control simulation with full CN and four soil/snow processes, and *OLD* – where the four soil/snow processes were excluded (Table 1). With respect to the observed permafrost $SOC_{0\text{-}1m}$ data based on NCSCD (472±34 PgC) [20] – the most extensive dataset of circumpolar permafrost carbon, *NEW* (461 PgC) is in much closer agreement while *OLD* (259 PgC) severely underestimates $SOC_{0\text{-}1m}$ (Figure 15). The incorporation of the specific soil/snow physical processes generally reduced grid-cell level differences with the NCSCD, along with other statistical improvements. Notably, too many grids in *OLD* contain very low (< 20 kgC/m^2) $SOC_{0\text{-}1m}$ density and few grid-cells accumulate 40 kgC/m^2 or higher; but in *NEW* the $SOC_{0\text{-}1m}$ density characteristics across the permafrost is improved and much better agreement with NCSCD.

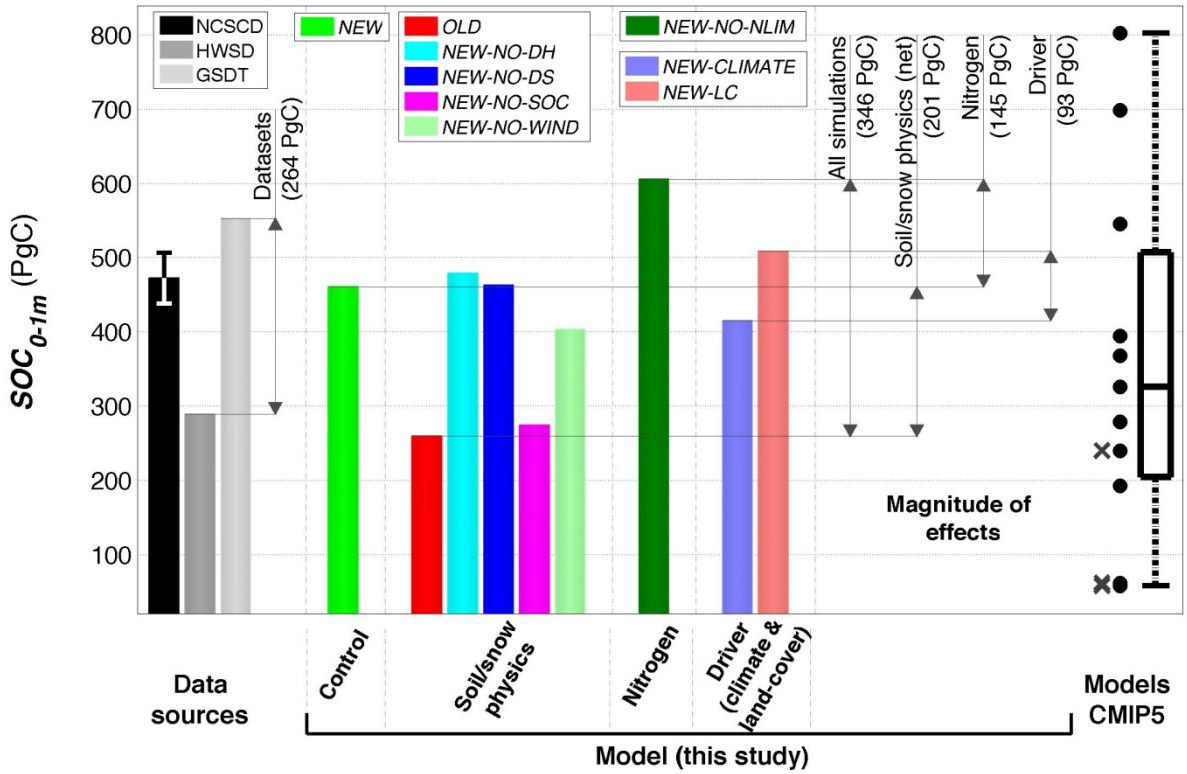


Figure 15: Total northern circumpolar permafrost $SOC_{0\text{-}1m}$ poleward of 45°N (using observed permafrost extent from the NCSCD data). Shown are estimates from three datasets for SOC encompassing the entire circumpolar permafrost (NCSCD, HWSD, GSDT), nine simulations from the current study (Table 1), and CMIP5 models. *NEW* ("Control" simulation) is a standard version of the model, containing all four soil/snow physical processes, fully prognostic CN cycle. "Soil/snow physics" differences from *NEW* were tested using four simulations each excluding one of four newly implemented soils/snow processes as a time from *NEW* (i.e., *NEW-NO-DH*: no depth hoar, *NEW-NO-DS:* no deep soil, *NEW-NO-SOC*: no SOC impacts on soil properties, *NEW-NO-WIND*: no wind compaction of snow), and a simulation that lacks all these processes (*OLD*). One simulation ("Nitrogen") was designed where N-limitation was relaxed from *NEW* (*NEW-NO-NLIM*). These seven aforementioned simulations were driven using the CRUNCEP reanalyses and land-cover data (LC1). Two simulations ("Driver") were performed to quantify the driver

uncertainties: *NEW-CLIMATE* (using NCEP/NCAR reanalysis instead of CRUNCEP in *NEW*), and *NEW-LC* (using LC2 land-cover instead of LC1 in *NEW*). For the CMIP5 models, black circular dots represent individual models, and the corresponding box plot represents the minimum, $25^{th}/50^{th}/75^{th}$ percentiles, and the maximum. The CMIP5 models with nitrogen cycle (amongst those with the lowest $SOC_{0\text{-}1m}$) are also denoted with cross marks alongside.

Ecologically, the SOC accumulation depends on mutual interactions of abiotic (climatic) and biotic factors, such as soil temperature, soil moisture content, litter amount and quality, N availability, and other factors affecting microbial decomposers in carbon pools. With the newly added cold-region soil/snow processes, results show that the simulated SOC gain is an outcome of two competing processes. *(1)* Decrease in total litter input (leaves + roots) in *NEW* vs. *OLD* in most areas of the permafrost – this causes lower accumulation of SOC due to diminished source of carbon to soils. The lower litter production in *NEW* is due to the cooler/drier soils and a slightly stronger N-limitation on photosynthesis in *NEW*, all of which impede gross primary production (GPP) and net primary production (NPP) by comparable amounts. *(2)* Reduced litter and soil decomposition rates (i.e., heterotrophic respiration) induced by cooler and drier soils in NEW than for decomposition in *OLD* – this favors more SOC accumulation. Overall, the temperature and moisture induced stresses on heterotrophic respiration so strongly compensates for the reduced litter inputs that the net carbon flux (towards land) remains higher in *NEW* during the entire buildup process of SOC, and the accumulation also continues for much longer timescales. Consequently, driven primarily by changes in soil temperature/moisture rather than litter the SOC equilibrates to substantially higher steady-state values throughout the permafrost areas. The tundra-dominant areas benefit the most, gaining $SOC_{0\text{-}1m}$ by 131 PgC (~2.45 times of OLD amount), followed by 42 PgC increase in tree-dominant areas (mostly from boreal forests), and 27 PgC in other areas. In the remainder of this study we present the physical mechanisms of such changes in the NHL SOC due to the individual cold-region soil/snow physics process, and compare the magnitude of their importance with that from N-cycle, and driver-induced uncertainties from climate and land-cover datasets.

Individually, the soil/snow processes are counterbalancing the SOC amount. The SOC-driven impacts on soil physical properties, and wind compaction of snow contribute to strong increases in simulated SOC, while depth hoar has a significant negative contribution to the overall SOC increase in *NEW* (Figure 16a). These changes are driven by relative magnitudes of impacts on soil temperature and moisture from these individual processes (Figure 16a). Generally, processes that cool and dry the soils (in mean annual scale) increase SOC accumulation. SOC-induced soil properties result in strongest simultaneous decreases in soil temperature and moisture, as the $SOC_{0\text{-}1m}$ across the NHL permafrost is rich in SOC. Insulation characteristics of SOC keeps the mean annual soil temperature low and increases the porosity and the hydraulic conductivity which produce drier soils. Notably, the top soils (0—30 cm soil) are often the driest due to the highest SOC densities according to the NCSCD data. With predominantly shallow root depths of the permafrost vegetation, the simulated evapotranspiration in *NEW* also decreases – additionally impacting biogeochemical parameterizations in the model that are dependent on it. In contrast to the perennial impacts of SOC on soil physical properties, wind compaction of snow impacts soil properties in winter – cooling soils due to reduced snow of insulation to outgoing ground heat. This effect primarily occurs in tundra and prairie regions, which are dominated by strong wind speeds[17]. Depth hoar effect is also seasonal, warming the soils during peak winter through stronger insulation. This effect is restricted to taiga (boreal) region where the snow is thick

enough to favor such crystal formation. In these cold-regions, any process that contribute to soil cooling can simultaneously limit the availability of soil water by favoring larger fraction of ice, and vice versa. Finally, inclusion of deep soil energetics nominally cool and dry the soils within 0—1 m, and its impact on $SOC_{0\text{-}1m}$ accumulation was found to be statistically insignificant. This is perhaps counterintuitive as deep soil energetics have been shown to play a strong role in increasing model-diagnosed near-surface permafrost area. Indeed our results are consistent with such an assessment, and exclusion of deep soils in our simulations contributes to the largest loss of model-diagnosed permafrost area than due any of the other three soil/snow processes. However, the key to our results and the previous studies is that near-surface permafrost is diagnosed using ~0—3 m of soil column (based on standard protocol). While incorporation of deep soils has strongest impact on soil temperatures at ~3 m, and this effect decreases towards the ground surface. Therefore specifically for soil biogeochemistry, we suggest that for current models that simulate carbon in the top 1 m of soil column, deep soil energetics are likely to have minimal impacts in estimating $SOC_{0\text{-}1m}$.

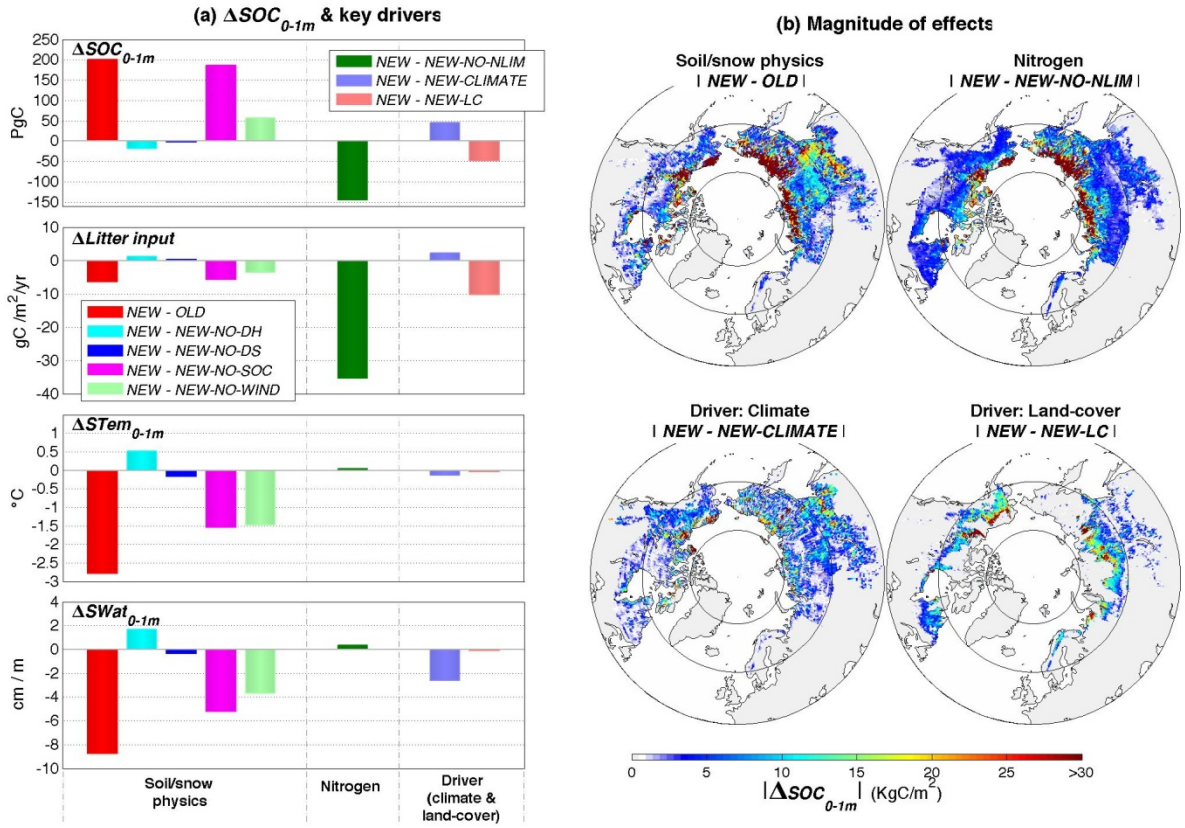


Figure 16: **(a)** Anomaly (Δ) in permafrost total $SOC_{0\text{-}1m}$ and annually averaged values in key biophysical drivers (*Litter input*, $STem_{0\text{-}1m}$: soil temperature in 0—1 m soil column, $SWat_{0\text{-}1m}$: soil water in 0—1 m soil column) in different simulations with respect to *NEW* (the control). All the values are using steady state estimates after completion of spin up, and using output from permafrost-affected areas only (i.e., the NCSCD grid cells). Except for the *NEW-NO-DS* (i.e., no deep soil in *NEW*), the results of $\Delta SOC_{0\text{-}1m}$ from other simulations were found to be statistically significant (confidence level > 99%). **(b)** Magnitude (in absolute terms) of net individual effects in modeling of permafrost $SOC_{0\text{-}1m}$, due to: soil/snow physics processes, nitrogen dynamics, and driver uncertainty from climate and land-cover data. For both magnitude and direction of these effects.

The relaxation of N-limitation in *NEW* – denoted by the simulation *NEW-NO-NLIM* – increases the permafrost $SOC_{0\text{-}1m}$ storage by 145 PgC (Figure 16a). This experiment represents a scenario where the soil is sufficiently rich in mineral N to completely facilitate photosynthetic assimilation (aboveground impact) as well as microbial decomposition (belowground impact). While plants only uptake from mineral N pool for photosynthetic assimilation, microbial decomposition is dependent on both mineral and organic N pools. In NEW results, the soil microbial decomposition is not N limited, but the aboveground processes (i.e., GPP, NPP, and litter production) are. This leads to a strong increase in litter production and higher $SOC_{0\text{-}1m}$ in *NEW-NO-NLIM* as compared to *NEW*.

With decomposition not being N-limited, the aforementioned 145 PgC represents the maximum reduction of permafrost $SOC_{0\text{-}1m}$ due to N limiting reduced above ground biomass in the current modeling framework. Yet, this amount (145 PgC) in absolute terms is smaller than that due to the four chosen soil/snow processes (201 PgC). Geographically too, more permafrost areas are more strongly impacted by the cold-region soil/snow physics than that due to N-cycle (Figure 16b). Notably, the $SOC_{0\text{-}1m}$ gains in *NEW-NO-NLIM* as compared to *NEW* is the strongest in the tundra-dominant grid-cells, followed by the other herbaceous grid-cells, and then the tree-dominant grid-cells. Nonetheless, these are smaller than corresponding increases in SOC due to inclusion of the soil/snow processes alone. Given that many cold-region biogeophysical processes are not represented in current generation of land surface models, their importance can only be higher. Finally, climatic and the land-cover driven uncertainties (from *NEW-CLIMATE* and *NEW-LC*, respectively; see Figure 15) result in $SOC_{0\text{-}1m}$ anomalies of ~-48 PgC (*NEW* minus *NEW-CLIMATE)* and 46 PgC (*NEW* minus *NEW-LC)* around *NEW*, respectively. Such differences, accounting for ~±10% uncertainty in $SOC_{0\text{-}1m}$ around *NEW*, strongly establish model-drivers as another important source of uncertainty in permafrost regions. In *NEW-CLIMTE* (driven by NCEP/NCAR), climatic variable values such as for precipitation and atmospheric humidity are higher than in *NEW* (driven by CRUNCEP) over much of the permafrost regions. This increases the soil moisture and favors more decomposition in *NEW-CLIMTE*, and reduces the $SOC_{0\text{-}1m}$ (Figure 16a). In *NEW-LC* (driven by LC2 land-cover data, boreal forests occupy more area (by ~2.67 million-km^2) compensated primarily by lower tundra extent. In such forested regions, the associated increases in litter production and input into soils drives the higher accumulation of $SOC_{0\text{-}1m}$ than in *NEW*. In most regions where forest vs. tundra difference occurs, the land-cover often becomes the dominant source of uncertainty greater than those from even soil/snow differences (Figure 16b).

In a broader context, the significance of our results is as follows. *(1)* We highlight a strong role of cold-region soil/snow processes in the estimation of permafrost SOC using a TBM, a coupled model accounting for biogeophysical and biogeochemical processes. While continued development of more sophisticated TBM is an ongoing pursuit, accurately representing cold-region biogeophysical processes in models become rather more critical as the drivers of biogeochemical processes themselves. Therefore, models that currently lack implementations of cold-region specific soil/snow processes are likely to benefit by including them in their TBM schemes. There is however a drawback due to substantially higher computational costs associated with longer spin-up timescales from such processes, in particular due to slow response of tundra ecosystems. *(2)* A combination of differences from soil/snow biogeophysical

processes, N-cycle, and driver-induced (climate, and land-cover) uncertainties can cause a large range in permafrost SOC. In our nine simulations where we selectively vary these factors (Figure 15), the permafrost $SOC_{0\text{-}1m}$ varies from 260—606 PgC (from *OLD* to *NEW-NO-NLIM*). This range (346 PgC) is equal to 73% of the estimated total $SOC_{0\text{-}1m}$ as compared to the NCSCD, and is dominantly explained by the differences in soil/snow physical processes, followed by N-cycle. Further, this divergence increases to 231—652 PgC (range of 419 PgC, equal to 88% of NCSCD total estimate) if we include the contribution from driver related uncertainties on *OLD* & *NEW-NO-NLIM* (not encompassed in the listed simulations). Such differences are comparable to the multi-model divergence in permafrost $SOC_{0\text{-}1m}$ from contemporary projects like the CMIP5 (60—820 PgC), and the MsTMIP where largest uncertainty in SOC stocks amongst participating models occur in the 40—65^{o}N latitudes [10]. While it is not possible to isolate the contribution of individual factors on the divergent SOC estimates in multi-model studies, the use of a single modeling framework, as used in the current study, allows us to do so. Moreover, the results robustly point to the critical importance of biogeophysical process representations on model biogeochemistry, besides the role of N-cycle, and climate and land-cover drivers. *(3)* Substantial uncertainties still exist in the currently available SOC datasets for permafrost regions (Figure 15), which has implications for models that use a SOC dataset to constrain soil properties (e.g., the current study) rather than using the prognostically computed SOC pools to do so. Given that SOC-driven soil physical properties have the singularly most important impact on the computation of SOC itself, the latter approach would be beneficial in model projections to include the feedbacks of large-scale SOC reduction on soil properties. However the former has advantages for calculating current day permafrost SOC, because it can constrain the soil properties more accurately. But this depends on the SOC dataset used. While the NCSCD is reasonably accurate over the permafrost, the limited choice between HWSD and GSDT as the only global datasets in offering introduces challenges. *(4)* Uncertainty in SOC data also affects rigorous model evaluation. Our analysis for the Alaskan permafrost (Figure 17) corroborates this issue. Here, confronting the NCSCD with an alternate dataset of Alaskan $SOC_{0\text{-}1m}$ shows a 17 PgC divergence between the datasets (~116% of total NCSCD estimates), which is greater than simulated differences due to soil/snow physics (12 PgC), as well as the combined contributions from climate and land-cover uncertainty (16 PgC). Nonetheless, these SOC datasets are still very valuable to qualitatively evaluate modeling improvements in biogeophysics from the incorporated soil/snow processes. For example, while *NEW* can simulate the high SOC densities (exceeding 50 kgC/m^2) in the northern Alaskan regions (present in both the datasets), corresponding estimates in *OLD* remain very low. *(5)* Finally, our results have important implications on studies of future permafrost SOC degradation. In response to climate and soil warming, the SOC balance in the entire NHLs will be determined by net impacts of two opposing mechanisms: enhanced soil N-mineralization leading to higher vegetation growth and litter productivity (negative feedback), and increased SOC decomposition to atmosphere (positive feedback). Using the cold-region soil/snow processes as the case study, here we find that while warmer soils in *OLD* only nominally increase GPP, NPP and litter production vis-à-vis in *NEW* (~2—3% change in these fluxes between the two simulations), the corresponding SOC reductions from soil warming/drying (by 201 PgC) are staggering. Therefore, we argue that the link between warming induced gains in permafrost SOC stocks from enhanced N-mineralization and fertilization of plant growth in the future will be much weaker than the offsetting decreases from increased soil decomposition. This is broadly in agreement with a similar inference from a recent study on climate-driven degradation of permafrost SOC. We

conclude by suggesting that biophysical feedbacks of ecosystem change are most likely to dominate the response of permafrost SOC in the future, which reinforces the importance of accurately accounting for such processes as drivers of biogeochemical processes themselves in the earth system modeling framework.

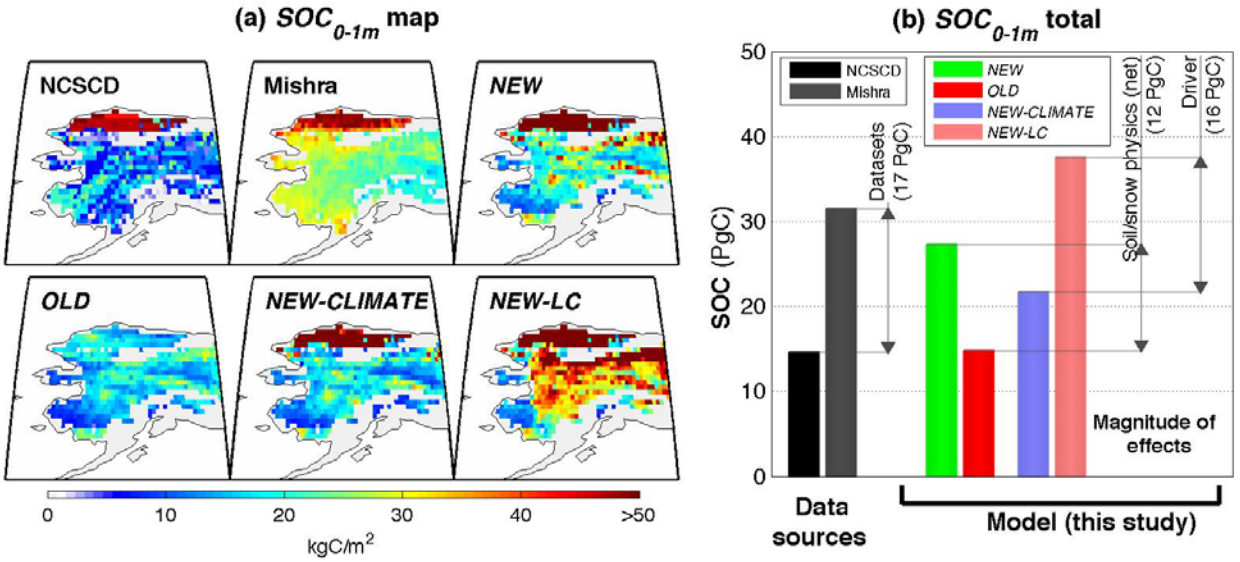


Figure 17: Alaskan permafrost $SOC_{0\text{-}1m}$. **(a)** Geographical distribution in two selected datasets (NCSCD, and four simulations (*NEW, OLD, NEW-CLIMATE, NEW-LC*) to highlight differences due to soil/snow physics processes, and driver–induced uncertainties. **(b)** Corresponding total estimates. Notably at this regional scale, the land-cover and climate can yield stronger differences than from soil/snow physics in both spatial and total estimates.

1.6. CESM-ISAM Coupling (Barman et al., 2011)

We coupled the current version of ISAM with the Community ESM (CESM1), to develop a flexible ESM modeling framework: CESM-ISAM (Figure 18). The CESM-ISAM retains the existing LSM in CESM, i.e., the Community Land Model 4 (CLM4), and allows both the ISAM and CLM to choose from all of the existing configurations available in CESM. Additionally, the resulting framework has been designed to incorporate multiple LSMs into the CESM, by adopting a flexible approach to coupling through the flux coupler (Figure 18). The purpose of this general modeling framework is to carry out equivalent climate simulations using multiple LSMs with the rest of the component models being the same, allowing a direct comparison of the effects of different land surface representations on corresponding feedbacks to climate change. Such a modeling framework establishes multiple opportunities to investigate the role of varying representation of land surface processes (such as from biogeophysics and biogeochemistry) on coupled land-atmosphere interactions.

CESM-ISAM Coupling Approach (using the CESM Flux coupler)

Figure 18: ISAM-CESM coupling flowchart. For details, see notes along with the Figure caption below.

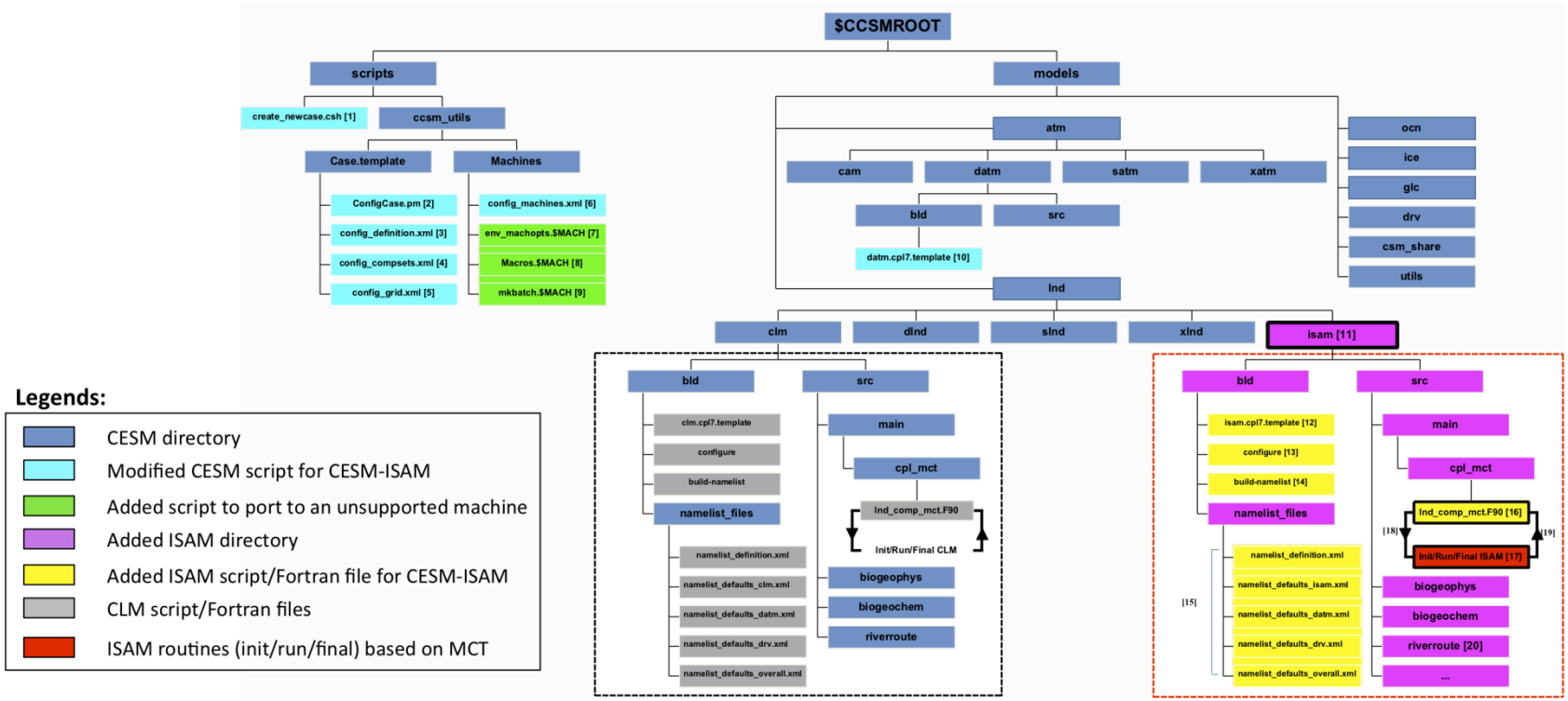
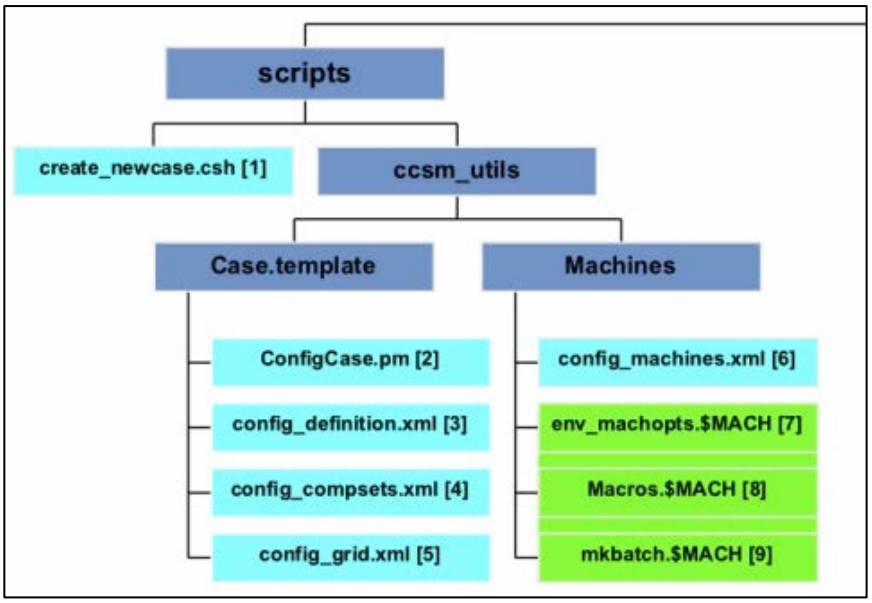


Figure 18 (continued) ISAM-CESM coupling flowchart. Shown is a *part* of the full flowchart from above, with relevant notes of the coupling process (Steps: 1—9)



1. Set ISAM path (**models/lnd/isam/bld/isam.cpl7.template**) in CESM as an alternate land model

2. Define ISAM as a new component with other existing components in the CESM framework

3. Add ISAM as an alternate land component; Define a new namelist group & options for ISAM

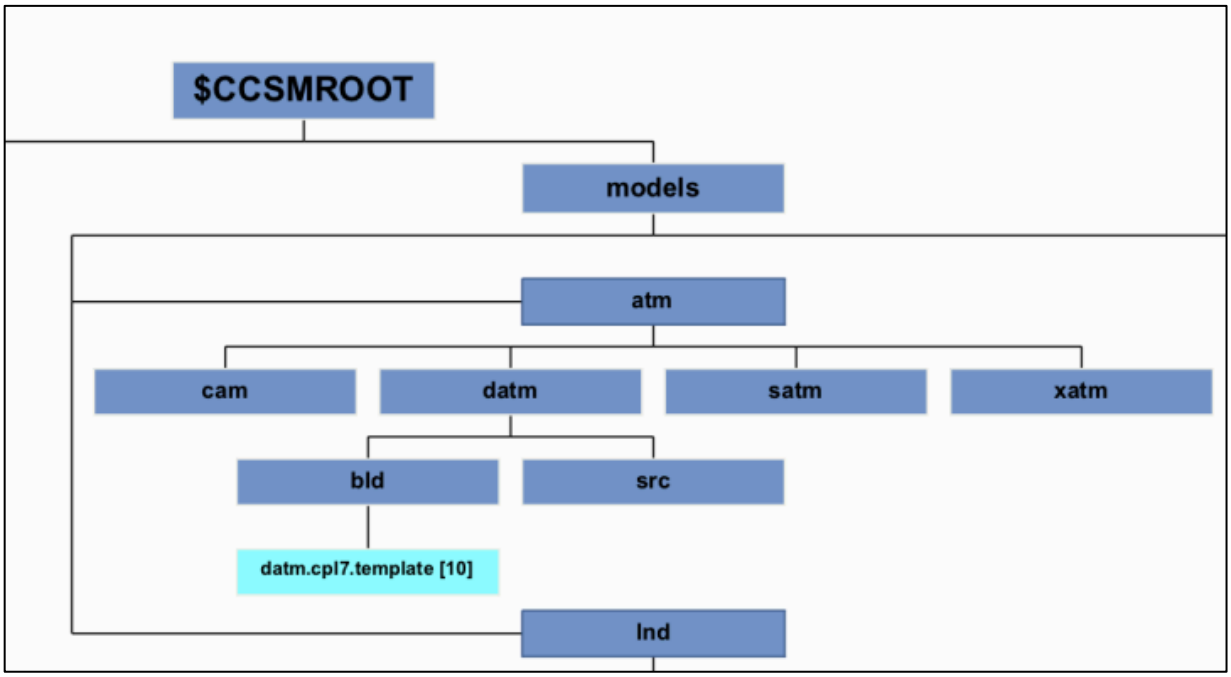
4. Define/Add new component sets and configurations, replacing CLM with ISAM as the land component (e.g., I_isam, F_isam, B_isam corresponding to I, F and B 'compsets' respectively)

5. Define/Add new ISAM grids (e.g., $0.5^{o}\times0.5^{o}$); Land-atmosphere mapping files for corresponding ISAM grids are generated offline using the SCRIP package

6. In an unsupported machine ($MACH), add machine settings for porting CESM/CESM-ISAM

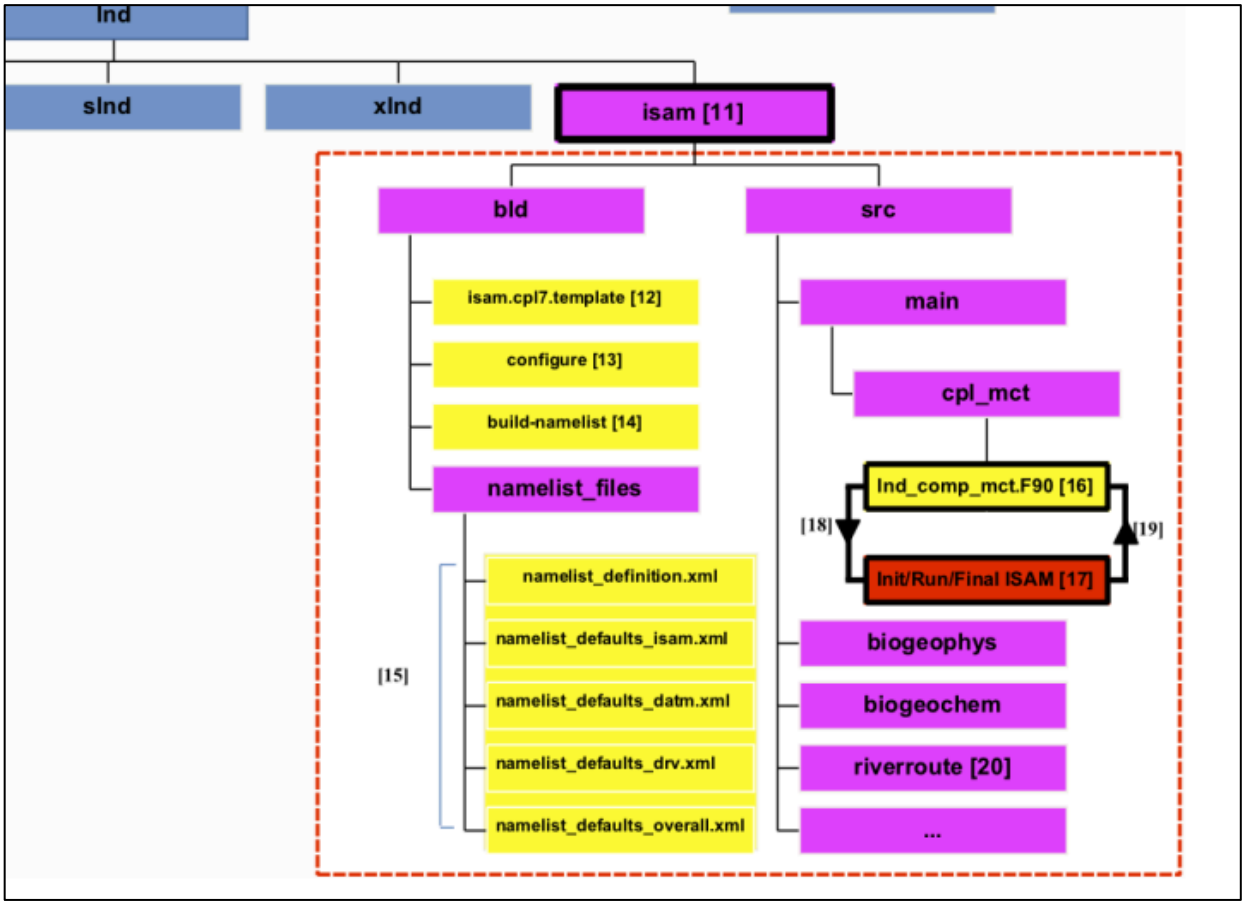
7-9. Required files for porting to a new, unsupported machine (See the CESM1 User's Guide)

Figure 18 (continued) ISAM-CESM coupling flowchart. Shown is a *part* of the full flowchart from above, with relevant notes of the coupling process (Steps: 10)



10. Add support for new ISAM grid(s) for atmospheric data (DATM) driven 'compsets'

Figure 18 (continued) ISAM-CESM coupling flowchart. Shown is a *part* of the full flowchart from above, with relevant notes of the coupling process (Steps: 11—20)



11. ISAM land model root directory in CESM-ISAM (Corresponding CLM source code hierarchy is also shown in the flowchart for comparison with the ISAM counterpart)

12. Generates three required scripts for building ISAM in CESM-ISAM analogous to the three scripts generated for CLM (isam.buildexe.csh, isam.buildnml.csh, isam.input_data_list)

13. Add available paths ("Filepath") for ISAM source directories

14. Builds a land model namelist for the defined CESM configuration, which contains CESM-specific control parameters; ISAM-specific namelist options are read using another namelist

15. Define and assign default values of the land model namelist options in CESM

16. The main interface between the CESM driver/coupler and ISAM; adapted from the corresponding MCT based CLM module (clm/src/main/cpl_mct/lnd_comp_mct.F90)

17. ISAM initialization/run/finalization methods; initializes SPMD, global segmentation map, land Domain; imports atmospheric inputs from the coupler to the land, runs the land model, and exports output back to the coupler

18-19. Fluxes/States from the coupler to the land and from the land to the coupler, respectively

20. The River Routing Model (RTM), extensively modified for ISAM data structures/grids from the original CLM version

1.7 ISAM model applications and comparing its results with other LSM model results

Improvement of ISAM LSM, to reduce uncertainties in estimated carbon and nitrogen fluxes, and understanding of how Earth's climate, biogeochemical systems, and human activities

interact over the 21st century and beyond are the overarching objectives of this project. Improving the representation of the biophysical and biogeochemical processes in LSM models requires extensive comparison of model results with observations. This process is difficult and time intensive. Past data-model and model-model intercomparisons have strengthened the representation of key processes in ISAM.

(1) The Multi-Year FACE experiment project is focusing on outputs from the Duke University and Oak Ridge National Laboratory multi-year FACE experiments. This National Center for Ecological Analysis and Synthesis (NCEAS) project is being headed by Dr. Richard Norby of ORNL**.** Unlike previous model-data intercomparison projects which focused on models' capacities to capture short-term and temporally integrated C and water, the current intercomparison focuses predominantly on multi-year trends in measured processes of the C, water, and N cycles of the experimental forest stands and their responses to elevated CO_2 concentrations.

(2) *Multi-Scale Synthesis and Intercomparison Project (MsTMIP)* is a formal model intercomparison activity with standardized model simulations within an integrated evaluation framework. This activity builds off current and past North American Carbon Program (NACP) syntheses activities, with an overall goal of providing feedback to the terrestrial biospheric modeling community to improve the diagnosis and attribution of carbon sources and sinks across regional and global scales. As part of this effort, we publish following publications:

(3) *TRENDY* project is a part of a consortium of Dynamic Global Vegetation Model (DGVM) groups set up a project to investigate further the spatial trends in Net Biome Production (NBP) and agreed to perform a factorial set of DGVM simulations over the historical period, 1901 - 2014.

As part of these projects, my group members have performed a series of model sensitivity simulations in order to understand the partitioning of the observed net ecosystem exchange (NEE) among processes such as climate variability, CO2 fertilization, nitrogen limitation, current land management, and the recovery from historical land use and disturbance. In collaboration with other researchers participating in these projects, we participated on these data-model intercomparision projects to further evaluate and improve our ISAM model performance at the global scale. In addition, we published a series of research articles outlining the approach we used to evaluate land ecosystem models as well as the findings of the data-model comparison results analyses (De Kauwe et al., 2013, 2014; Le Quere et al., 2013, 2014, 2015; Christoffersen et al., 2014; Fisher et al., 2014; Walker et al., 2014; Zaehle et al., 2014; Zscheischler et al., 2014; Ahlström et al., 2015; Mao et al., 2015; Schwalm et al., 2015; Smith et al., 2015; Tian et al., 2015; Ito et al., 2016; Zhao et al., 2016).

2. Scaling the ISAM Land Surface Model through Parallelization of Inter-Component Data Transfer

Computational climate modelling is a key problem for contemporary high-performance computing. The numerous coupled phenomena that comprise the climate present a uniquely demanding challenge. Various elements have time scales that range from seconds and minutes to years and millennia, and space scales ranging from individual rows of farm fields to the entire globe. Given the limits of interest and expertise, models focusing on just a few effects

are typically developed in isolation by researchers specialized in each particular subarea. The interactions between different model components are handled either via on-line coupling or consumption of saved model output and processed sensor data from 'upstream' components as input to 'downstream' components. The former can create substantial computational challenges, while the latter can require storage of and access to massive data sets. In this study, we explore some elements of the off-line case in the Integrated Science Assessment Model (ISAM). The ISAM code presents two different points for data transfer between components. The first inter-component data dependence in ISAM arises from the time-dependent climatic variables that serve as boundary conditions for its computations. In an online coupled simulation, such as with other components of CESM, these variables become available as they are computed, and may be dependent on results of previous land-surface time steps. In a stand-alone off-line simulation, these are provided in NetCDF files, generated and distributed by external groups. The second point of inter-component data transfer occurs between separate runs performing different stages of model spin-up. In order to set initial conditions for each of these runs, large portions of model state from the preceding run must be preserved. This study also describes the other changes necessary to enable ISAM to scale to large supercomputer systems. We also address the load balancing issues in distributing model grid points among processors in the parallel system. The end result of this study is that the model is able to strong-scale a whole-Earth land surface simulation from 256 to 16k processors of NERSC's Edison Cray XC30 system with over 50% parallel efficiency, and with a clear path to further improvement.

The various points at which a processor must wait for other processors to coordinate a synchronous global operation present opportunities for load imbalance to negatively impact performance. At these points, the waiting time is determined by the execution time of the most heavily loaded processor. The consequent loss of performance can be mitigated by combinations of improving load balance and removing the need for synchronization. We can measure the overall cost to performance using the imbalance time metric. All 'Idle' time displayed in our figures is attributable to imbalance time, since there is no operation-dependent communication latency or other source of underutilization. There are several distinct sources of potential load imbalance in ISAM. The different sets of points assigned to different processors may generate different cumulative loads over various time scales. Some operations may be performed serially on a single processor while others wait. Contention for shared resources may induce imbalance where it is not otherwise inherent in the operation being performed. There are also synchronous operations that can be reduced. They can be made less frequent directly and by having the code do more with each one.

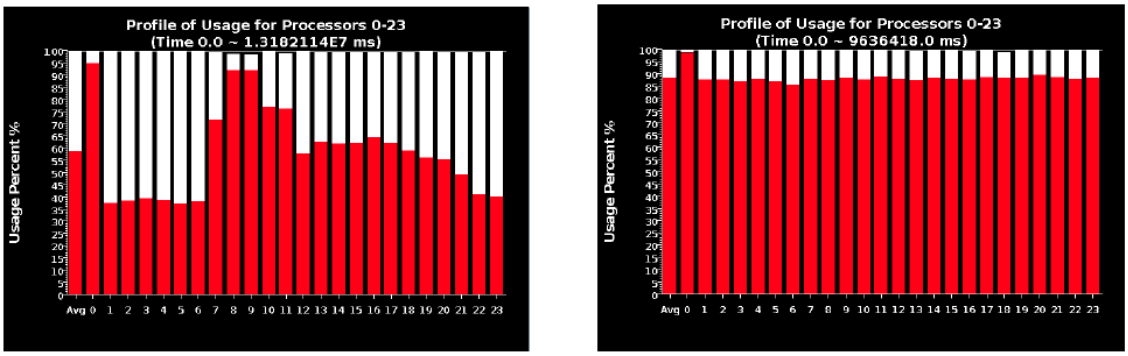
Experimental Setup. Our scaling experiments reported were run on the Cray supercomputers hosted at NERSC, Hopper and Edison. Hopper is a Cray XE6 with 6,384 nodes each containing a pair of 12-core AMD 'Magny Cours' processors running at 2.1 GHz. Edison is a Cray XC30 with 5,576 nodes each containing a pair of 12-core Intel 'Ivy Bridge' processors running at 2.4 GHz. Edison's processors support 2-way HyperThreading SMT, but our experiments were run with a single thread per core. Both systems are served by Lustre parallel filesystems, on which all files relevant to our experiments were stored. On both machines, we used the Intel Fortran Compiler. Binaries for runs on Hopper were compiled with version 13.1 and binaries for runs on Edison were compiled with version 14.0. All of our figures show the

best results on each scale on each machine. Due to run-to-run variability (from node placement and contention with other jobs for network and filesystem resources), some experimental runs were up to 16% slower than the speeds displayed.

Computational Load Balance. In initial versions of ISAM, land surface points were mapped to MPI ranks in a uniform blocked fashion: the total number of points being simulated was divided by the number of ranks in the job, and each rank took responsibility for a consecutive chunk of points. As generated from the data set representing which points on Earth's surface were land (rather than open water), the points were ordered along successive lines of latitude. Each processors was thus likely to receive points that were geographically nearby and thus both similar in systematic spatial work characteristics and closely correlated in temporal cycles and climate variations.

The result of this structure was that entire regions of the planet presenting high workload would be assigned to one set of processors, while regions presenting much less work would be assigned to others. The consequent aggregate utilization of each processor was highly varied. This can be seen in Figure 19a, generated from traces of a 24-processor run in the Projections tool. Since the code incurs regular global synchronization, even short-term dynamic load variation causes performance to suffer.

To address this issue, the code was adapted to distribute the points cyclically, in a round-robin fashion. The new mapping is illustrated in Figure 20. This ensures that the points of any given region are spread across many processors, and that each processor hosts points from distinct regions. This results in mixing high-load points with low-load points, such that they tend to average out. The processor utilization after switching to this mapping can be seen in Figure 19b.



(a) Before: block mapping (b) After: round-robin mapping

Figure 19. Processor utilization of each rank visualized in the Projections tool.

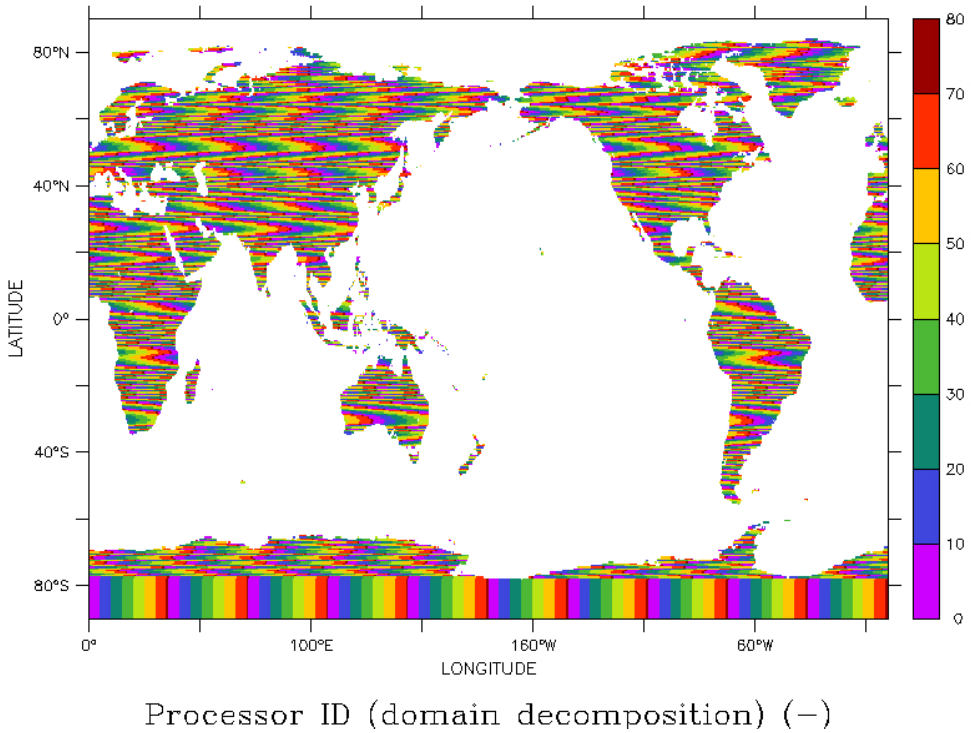


Figure 20: Round-robin mapping of land surface grid points across 72 processors. Process rank is represented by color of each point.

Applying this optimization enabled strong scaling from a single node to a few dozen nodes as in a small cluster.

Parallel Output to Reduce Memory Footprint. The ISAM biogeochemistry spinup is parameterized by preliminary steady-state values of the biogeophysics portion of the model. This requires saving the values of the coupling variables at every simulated point over a sufficient span of time steps to provide accurate results. A similar data set is also necessary to checkpoint the biogeophysics component's state, so that it can be restarted after system failures or job length expiration. The largest of these elements are a set of 10 variables that are recorded for each of the 28 PFTs at every point with a weekly resolution over the course of a simulated year. At a resolution of 360 x 720 (i.e. 0:5^{o}) and recorded in double precision, each variable occupies 2.8 gigabytes.

This demand for large-memory nodes is problematic as discussed in Miller et al. (2014). By parallelizing the output of these variables across multiple nodes, we reduce the maximum per-node memory footprint of the output operation. A reduced footprint thus enables execution on more plentiful computational resources and more rapid scheduling of jobs on those resources. On Hopper, using version 1.2.0 of Parallel netCDF, we were able to write the 25 GB making up a complete set of these largest variables to the Lustre 'scratch' filesystem in 17 seconds, for an overall bandwidth of 1:46 GB/s.

Input Optimization. The basic structure of the input process involves three steps. A process reads the appropriate point in the time series provided by the input files, using the NetCDF library. It then computes the spatial interpolation to the simulated land surface points. Finally, the

interpolated data are distributed to the processes according to which grid points they are responsible for, as described in Miller et al. (2014). The initial design of this input presented many impediments to scalability. In the remainder of this section, we discuss how these limitations have been largely eliminated.

First we modified ISAM to reuse already-prepared data, we improved performance by a factor of 1:7x on 1024 ranks of Hopper and 1:2x on 1024 ranks of Edison. This improvement comes from both reduced time spent performing collectives, and reduced imbalance time waiting on heavily-loaded processors to reach each collective. On 1024 processors of Hopper, where the original and first optimized version of the code obtain their best performance, the optimized code spends 88% less CPU time performing collectives and 24% less CPU time idling. The decline in collective time accounts for 62% of the 1:7x speedup and the decline in idle time accounts for a further 34% of the speedup.

On 1024 processors of Edison, where the original and first optimized version of the code also obtain their best performance, the optimized code spends 30% less CPU time performing collectives and 18% less CPU time idling. The decline in collective time accounts for 64% of the 1:2x speedup and the decline in idle time accounts for a further 33% of the speedup.

With input data read from the filesystem every few steps, the time per step scales poorly due to an Amdahl's law bottleneck on the time to access the filesystem and interpolate the data. Additionally, contemporary supercomputers offer high-band width parallel filesystems to support their computational capabilities. By reading input data in only a single rank, ISAM was limited to the bandwidth of a single node. Thus, our next optimization to ISAM's input process was to read and interpolate many steps worth of input data in parallel. At model time steps where data must be read, each process reads and interpolates data for a step computed by incrementing the current time step by its rank. At each subsequent step, the responsibility for distributing data cycles across the ranks until every rank has served as root once.

In theory, this can reduce the elapsed wall time spent on reading and interpolation by O(P), since P such steps are performed in parallel. This is potentially limited by available bandwidth both in accessing the file data from the filesystem and in interpolating it in memory. At larger scales, we observe this effect. The improvement provided by this optimization is 2:76 x on 1024 processors of Hopper and 1:3x on 1024 processors of Edison. In both cases, the reduction in idle time accounts for the bulk of the improvement. On both systems, this optimization allows the code to continue to gain performance at scales up to 2k processors, with efficiencies of 39% and 34% respectively, relative to the 256 processor baseline.

Having minimized idle time by fully parallelizing the reading and interpolation steps, the largest non-work portion of the execution time at the scaling limit of the code is spent in collectives. On 2048 processors, these consume 46% of CPU seconds on Hopper and 57% of CPU seconds on Edison. On both systems, the increases in collective times account for the bulk of increased time relative to runs on 1024 processors. To overcome this impediment, we observe that at the
first scatter operation after climate forcing data is read and interpolated, the P processors each have data available for an upcoming time step. However, in each scatter, only the cyclically selected root processor actually provides it. This misses a substantial opportunity for increased

parallelism in usage of network resources. We take advantage of this opportunity by converting the per-step MPI_Scatterv operation to an MPI_Alltoallv operation performed every P time steps. Rather than spatially scattering data representing the climate forcing at a singlepoint in time, we now transpose the data from its provided temporal distribution (each processor sends a distinct time step for every point) to a spatial distribution (each processor receives the time series for the points it owns). Once this is done, each processor can independently execute P time steps with no communication.

At first glance, this pre-distribution of input data may seem to dramatically increase memory usage on every processor. However, this is not the case. To see why, we first observe that the additional memory consumption is a constant, independent of P. Suppose there are *n* points in total, and each one requires *b* bytes of memory for a single time step's climate data. Each processor is responsible for n=P of those points. The data each processor reads from disk is *bn*. In the transposition, each processor receives the *bn=P* bytes for one future time step from each of the P processors. Thus, the total received data is just *bn* – exactly as much as every processor read from disk. For the NCEPQ climate data set, *b* = 24 and n = 192 x 94 = 18,048, totaling 423 kilobytes. For the CRU NCEP data set, b = 32 and n = 720 x 360 = 259,200, totaling 8 megabytes.

The effects of this optimization are striking. Where previously, roughly half the execution time was spent in collectives at just 2k processors, this optimization reduces that time to less than 1% on both systems. Additionally, idle times also fell by over 50% on both systems, due to the longer period between synchronization points and greater opportunity for dynamic load variation to average out. Moreover, read times (though representing only a small proportion of execution) also fell substantially because of this optimization. We conjecture that this decrease is due to reduced contention when accessing the filesystem, since different processors can reach this phase spread out by the load variation of P steps worth of work rather than just 1. Overall, this provides a 2:4x speedup on 2k processors of Hopper, and a 2:9x speedup on 2k processors of Edison. It also allows us to scale with continued speedups to 16k processors. In summary, from our baseline code, we have obtained speedups of 6:58x on 1024 processors of Hopper and 2:78x on Edison. With all of the optimizations applied, we strong scale from 256 processors to 2048 process with an efficiency of 88% on Hopper and 91% on Edison.

Conclusions. As originally proposed, we have accomplished the progression of developments necessary to scale the ISAM land surface model from single nodes and small clusters with unusually large per-node memory to much larger systems with more common configurations. These efforts included load balancing, conventional library-based output parallelization to reduce memory load, and parallel in-time data input. On Hopper, the result was strong scaling from 256 processors to 16k processors for a speedup of 21x, giving an efficiency of 32.9%. On Edison, the code exhibits a strong-scaling speedup from 256 processors to 16k processors of 32:9x, for an efficiency of 51.4%. These large-scale gains, and the associated performance increases at smaller scale, will enable greater scientific productivity for the users of ISAM and open the possibilities of increased resolution in time and space and greater physical fidelity for the simulated processes while remaining computationally feasible.

Research papers, Ph.D. Thesis, Presentations that have been sponsored completely or in part by DOE program

Research Papers (Total 27 Research Articles)

Schaefer, K. M., C. R. Schwalm, C. A. Williams, M. Arain, A. G. Barr, J. M. Chen, K. J. Davis, D. D. Dimitrov, T. W. Hilton, D. Y. Hollinger, E. Humphreys, B. Poulter, B. Raczka, A. D. Richardson, A. Sahoo, P. E. Thornton, R. Vargas, H. Verbeeck, R. S. Anderson, I. Baker, T. Black, P. V. Bolstad, J. Chen, P. Curtis, A. R. Desai, M. Dietze, D. Dragoni, C. Gough, R. F. Grant, L. Gu, A. Jain, C. J. Kucharik, B. E. Law, S. Liu, E. Lokipitiya, H. Margolis, R. Matamala, J. McCaughey, R. K. Monson, J. Munger, W. C. Oechel, C. Peng, D. T. Price, D. M. Ricciuto, W. J. Riley, N. T. Roulet, H. Tian, C. Tonitto, M. S. Torn, E. Weng, and X. Zhou (2012): A model-data comparison of Gross Primary Productivity: Results from the North American Carbon Program site synthesis, *J. Geophys. Res*., 117, G03010, doi:10.1029/2012JG001960.

Jain, A.K., P. Meiyappan, Y. Song, J. House, 2013: CO_2 emissions from land-use change affected more by nitrogen cycle, than by the choice of land-cover data, *Global Change Biology*, doi: 10.1111/gcb.12207.

Kauwe, M. G. De, B. E. Medlyn, S. Zaehle, M. Dietze, T. Hickler, A. Jain, Y. Luo, W. Parton, C. Prentice, P. Thornton, A. Walker, S. Wang, W.P. Wang, D. Warlind, E. Wang, K. Crous, K. Ellsworth, 2013: Water use and water use efficiency at elevated CO_2: a model-data intercomparison at two contrasting temperate forest FACE sites, *Global Change Biology*, doi: 10.1111/gcb.12164.

Le Qu´er´e, C., R. J. Andres, T. Boden, T. Conway, R. A. Houghton, J. I. House, G. Marland, G. P. Peters, G. van der Werf, A. Ahlstr ¨om, R. M. Andrew, L. Bopp, J. G. Canadell, P. Ciais, S. C. Doney, C. Enright, P. Friedlingstein, C. Huntingford, A. K. Jain, C. Jourdain, E. Kato, R. Keeling, S. Levis, P. Levy, M. Lomas, B. Poulter, M. R. Raupach, J. Schwinger, S. Sitch, B. D. Stocker, N. Viovy, S. Zaehle, and N. Zeng, 2013: The global carbon budget 1959–2011, *Earth Syst. Sci. Data*, 5, 165–185.

Yang, X., W. Post, P. Thornton, A. K. Jain, 2013: The distribution of soil phosphorus in terrestrial equilibrium, *Biogeosciences*, 10, 2525–2537.

Barman, R. A.K. Jain, M. Liang, 2014a: Climate-driven uncertainties in terrestrial gross primary production: a site-level to global scale analysis, *Global Change Biology*, DOI: 10.1111/gcb.12474.

Barman, R. A.K. Jain, M. Liang, 2014b: Climate-driven uncertainties in terrestrial energy and water fluxes: a site-level to global scale analysis, *Global Change Biology*, DOI: 10.1111/gcb.12473.

Christoffersen, B.O., Restrepo-Coupe N, Arain MA; Baker IT, Cestaro BP, Ciais P, Fisher JB, Galbraith DR, Guan X, Gulden L, van den Hurk B, Ichii K, Imbuzeiro HM, Jain AK, Levine N, Miguez-Macho G, Poulter B, Roberti D, Sakaguchi K; Sahoo A, Schaefer K, Shi M, Verbeeck H, Yang ZL, Araújo AC, Kruijt B, Manzi AO, da Rocha HR, von Randow C, Muza MN, J Borak, MH Costa, GG de Gonçalves, X Zeng, Saleska SR, 2014: Mechanisms of water supply and vegetation demand govern the seasonality and magnitude of evapotranspiration in Amazonia and cerrado *Agricultural and Forest Meteorology*, 182–183, 145– 155.

De Kauwe MG, BE Medlyn, S Zaehle, AP Walker, MC Dietze, YP Wang, Y Luo, AK Jain, B El-Masri, T Hickler, D Warlind, W Ensheng, WJ Parton, PE Thornton, S Wang, IC Prentice, S Asao, B Smith, HR McCarthy, CM Iversen, PJ Hanson, JM Warren, R Oren and RJ Norby (2014) Where does the carbon go? A model–data intercomparison of vegetation carbon allocation and turnover processes at two temperate forest free-air CO2 enrichment sites, *New Phytologist*, doi: 10.1111/nph.12847.

Fisher, JB, M Sikka, WC Oechel, DN Huntzinger, JR Melton, CD Koven, A Ahlström, AM Arain, I Baker, JM Chen, P Ciais, C Davidson, M Dietze, B El-Masri, D Hayes, C Huntingford, AK Jain, PE Levy, MR Lomas, B Poulter, D Price, AK Sahoo, K Schaefer, H Tian, E Tomelleri, H Verbeeck, N Viovy, R Wania, N Zeng, and CE Miller (2014): Carbon cycle uncertainty in the Alaskan Arctic, Biogeosciences, 11, 4271-4288, doi:10.5194/bg-11-4271-2014.

Le Quéré C, GP Peters, RJ Andres, RM Andrew, T Boden, P Ciais, P Friedlingstein, RA Houghton, G Marland, R Moriarty, S Sitch, P Tans, A Arneth, A Arvanitis, DCE Bakker, L Bopp, JG Canadell, LP Chini, SC Doney, A Harper, I Harris, JI House, AK Jain, SD Jones, E Kato, RF Keeling, KK Goldewijk, A Körtzinger, C Koven, N Lefèvre, F Maignan, A Omar, T Ono , G-H Park, B. Pfeil, B Poulter, MR Raupach, P Regnier, C Rödenbeck, S Saito, J Schwinger, J Segschneider, BD Stocker, B Tilbrook, S van Heuven, N Viovy, R Wanninkhof, A Wiltshire, and S Zaehle (2014), Global carbon budget 2013, *Earth Syst. Sci. Data*, doi:10.5194/essd-6-1-2014.

Miller P., Robeson, M., El-Masri, B., Barman, R., Zheng, G., Jain, AK, Kale., L. (2014) Scaling the ISAM land surface model through parallelization of inter-component data transfer. *Proceeding of the International Conference of Parallel Processing*, 422-431. IEEE, 2014.

Walker AP, PJ Hanson, MG De Kauwe, BE Medlyn, S Zaehle, S Asao, M Dietze, MT Hickler, C Huntingford, CM Iversen, AK Jain, M Lomas, Y Luo, H McCarthy, WJ Parton, IC Prentice, PE Thornton, S Wang, YP Wang, D Warlind, E Weng, JM Warren, FI Woodward, R Oren, and RJ Norby (2014), Comprehensive ecosystem model-data synthesis using multiple data sets at two temperate forest free-air CO2 enrichment experiments: Model performance at ambient CO2 concentration, *J. Geophys. Res. Biogeosci.,* 119, doi:10.1002/2013JG002553.

Zaehle S, BE Medlyn, MG De Kauwe, AP Walker, MC Dietze, T Hickler, Y Luo, YP Wang, B El-Masri, P Thornton, AK Jain, S Wang, D Warlind, E Weng, W Parton, CM Iversen, A Gallet-Budynek, H McCarthy, A Finzi, PJ Hanson, IC Prentice, R Oren1 and RJ Norby 2014: Evaluation of 11 terrestrial carbon–nitrogen cycle models against observations from two temperate Free-Air CO2 Enrichment studies, *New Phytologist*, doi: 10.1111/nph.12697.

Zscheischler J, AM Michalak, C Schwalm, MD Mahecha, DN Huntzinger, M Reichstein, G Berthier, P Ciais, RB Cook, B El-Masri, M Huang, A Ito, AK Jain, A King, H Lei, C Lu, J Mao, S Peng, B Poulter, D Ricciuto, X Shi, B Tao, H Tian, N Viovy, W Wang, YWei, J Yang, and N Zeng (2014), Impact of large-scale climate extremes on biospheric carbon fluxes: An intercomparison based on MsTMIP data, *Global Biogeochem. Cycles*, *28*, doi:10.1002/2014GB004826.

Ahlström A, MR Raupach, G Schurgers, B Smith, A Arneth, M Jung, M Reichstein, JG Canadell, P Friedlingstein, AK Jain, E Kato, B Poulter, SS Benjamin, D Stocker, N Viovy,

YP Wang,16 A Wiltshire, S Zaehle, N Zeng (2015) The dominant role of semi-arid ecosystems in the trend and variability of the land CO2 sink, *Science*, 348, 895-899.

El-Masri B, S Shu, AK Jain (2015): Implementation of a dynamic rooting depth and phenology into a land surface model: Evaluation of carbon, water, and energy fluxes in high latitude ecosystems, *Agricultural and Forest Meteorology*, 211, 85-99.

Le Quéré C, R Moriarty, RM Andrew, JG Canadell, S Sitch, JI Korsbakken, P Friedlingstein, GP Peters, RJ Andres, TA Boden, RA Houghton, JI House, RF Keeling, P Tans, A Arneth, DCE Bakker, L Barbero , L Bopp, J Chang, F Chevallier, LP Chini, P Ciais, M Fader, R Feely, T Gkritzalis, I Harris, J Hauck, T Ilyina, AK Jain, E Kato, V Kitidis, K Klein Goldewijk, C Koven, P Landschützer, SK Lauvset, N Lefèvre, A Lenton, ID Lima, N Metzl, F Millero, DR Munro, A Murata, JEMS Nabel, S Nakaoka, Y Nojiri, K O'Brien, A Olsen, T Ono, FF Pérez, B Pfeil, D Pierrot, B Poulter, G Rehder, C Rödenbeck, S Saito, U Schuster, J Schwinger, R Séférian, T Steinhoff, BD Stocker, AJ Sutton, T Takahashi, B Tilbrook, IT van der Laan-Luijkx, GR van der Werf, S van Heuven, D Vandemark, N Viovy, A Wiltshire, S Zaehle, and N Zeng (2015), Global carbon budget 2015, *Earth Syst. Sci. Data*, DOI:10.5194/essd-7-349-2015

Mao J, W Fu, X Shi, DM Ricciuto, JB Fisher, E Dickinson, Y Wei, W Shem, S Piao, K Wang, CR Schwalm, H Tian, M Mu, A Arain, P Ciais, R Cook, YD, D Hayes, FM Hoffman, M Huang, S Huang, DN Huntzinger, Akihiko Ito, A Jain, AW King, H Lei, C Lu, AM Michalak, N Parazoo, C Peng, S Peng, B Poulter, K Schaefer, E Jafarov, PE Thornton, W Wang, N Zeng, Z Zeng, F Zhao, Q Zhu and Z Zhu (2015), Disentangling climatic and anthropogenic controls on global terrestrial evapotranspiration trends, *Environmental Research Letters*, 10, doi:10.1088/1748-9326/10/9/094008.

Meiyappan P, AK Jain, J House (2015), Increased influence of nitrogen limitation on CO_2 emissions from future land use and land-use change, *Global Biogeochemical Cycles*, 30, doi:10.1002/2015GB005086.

Schwalm CR, DN Huntinzger, JB Fisher, AM Michalak, K Bowman, P Cias, R Cook, B El-Masri, D Hayes, M Huang, A Ito, AK Jain, AW King, H Lei, J Liu, C Lu, J Mao, S Peng, B Poulter, D Ricciuto, K Schaefer, X Shi, B Tao, H Tian, W Wang, Y Wei, J Yang, N Zeng (2015) Toward "optimal" integration of terrestrial biosphere models, *Geophysical Research Letters*,. 42, 4418–4428, doi:10.1002/2015GL064002

Smith, P, JI. House, M Bustamante, J Sobocká, R Harper, G Pan, P West, J Clark, T Adhya, C Rumpe, K Paustian, P Kuikman, MF Cotrufo, JA Elliott, R McDowell, RI Griffiths, S Asakawa, A Bondeau, AK Jain, J Meersmans and TAM Pugh (2015), Global Change Pressures on Soils from Land Use and Management, *Global Change Biology*, doi:10.1111/gcb.13068.

Tian H, C Lu, J Yang, K Banger, DN Huntinzger, CR Schwalm, AM Michalak, R Cook, P Ciais, D Hayes, M Huang, A Ito, AK Jain, H Lei, J Mao, S Pan, WM Post, S Peng, B Poulter, W Ren, D Ricciuto, K Schaefer, X Shi, B Tao, W Wang, Y Wei, Q Yang, B Zhang, N Zeng (2015) Global Patterns and controls of soil organic carbon dynamics as simulated by multiple terrestrial biosphere models: current status and future directions, *Global Biogeochemical Cycles*, 29, doi:10.1002/2014GB005021.

Barman R and AK Jain (2016), Comparison of effects of cold-region soil/snow processes and the uncertainties from model forcing data on permafrost physical characteristics, *Journal of Advances in Modeling Earth Systems*, DOI: 10.1002/2015MS000504.

Ito A, M Inatomi, DN Huntzinger, C Schwalm, AM Michalak, R Cook, AW King, J Mao, Y Wei, WM Post, W Wang, MA Arain, M Huang, H Lei, H Tian, C Lu, J Yang, B Tao, A Jain, B Poulter, S Peng, P Ciais, JB Fisher, N Parazoo, K Schaefer, C Peng, N Zeng, F Zhao (2016), Decadal trends in the seasonal-cycle amplitude of terrestrial CO2 exchange resulting from the ensemble of terrestrial biosphere models, *Tellus B*, 68 (28968) doi:10.3402/tellusb.v68.28968.

Shao J, X Zhou, Y Luo, G Zhang, W Yan, J Li, B Li, L Dan, JB Fisher, Z Gao, Y He, D Huntzinger, **AK Jain**, J Mao, J Meng, AM Michalak, NC Parazoo, C Peng, B Poulter, CR Schwalm, X Shi, R Sun, F Tao, H Tian, Y Wei, N Zeng, Q Zhu, and W Zhu (2016), Uncertainty analysis of terrestrial net primary productivity and net biome productivity in China during 1901–2005, *J. Geophys. Res. Biogeosci.*, 121, doi:10.1002/2015JG003062. Zhao F, N Zeng, A Ito, G Asrar, P Friedlingstein, A Jain, E Kalnay, E Kato, CD Koven, B Poulter, R Rafique, S Sitch, S Shu, B Stocker, N Viovy, A Wiltshire, S Zaehle (2016), Role of CO2, climate and land use in regulating the seasonal amplitude increase of carbon fluxes in terrestrial ecosystems: a multimodel analysis, *Biogeosciences Discussion*, doi:10.5194/bg-2016-121, 2016.

Zhao F, N Zeng, A Ito, G Asrar, P Friedlingstein, **A Jain**, E Kalnay, E Kato, CD Koven, B Poulter, R Rafique, S Sitch, S Shu, B Stocker, N Viovy, A Wiltshire, S Zaehle (2016), Role of CO2, climate and land use in regulating the seasonal amplitude increase of carbon fluxes in terrestrial ecosystems: a multimodel analysis, *Biogeosciences*, 13, 5121–5137.

Ph.D. Thesis

Barman, Rahul: Impacts of model-data uncertainties in the investigation of biogeophysical-biogeochemical interactions in the terrestrial northern high-latitudes, Ph.D. Thesis, University of Illinois, Urbana-Champaign, *awarded in August 2014*.

Presentations (Total 21 presentations)

El-Masri, B. and A. K. Jain (2012): Modeling the above and below ground carbon and nitrogen stocks in northern high latitude terrestrial ecosystems, *American Geophysical Union (AGU) Fall meeting*, San Francisco, CA, December 3 – 7, 2012.

Walker A. P.; S. Zaehle; M. G. De Kauwe; M. C. Dietze; P. J. Hanson; T. Hickler; A. K. Jain; Y. Luo; H. R. McCarthy; B. E. Medlyn; W. J. Parton; P. E. Thornton; S. Wang; Y. Wang; D. Warlind; E. Weng; R. Oren; R. J. Norby, 2012: Climatic drivers of variability in the response of NPP to elevated CO2. A model-data comparison at two FACE sites in the south eastern US, *Fall 2012 American Geophysical Union (AGU) Meeting*, San Francisco, CA, 3-7 December 2012.

Jain, A.K. (2013) Modeling the Above and Below Ground C and N Stocks in Northern High Latitude Terrestrial Ecosystems, *CESM Land Model and Biogeochemistry Working Groups Meeting*, NCAR, Boulder, CO, February 20 – 22, 2013.

Yang, X., W.M. Post, A. Jain, P. Thornton (2013): Towards a representation of phosphorous dynamics in earth system models: Development of global scale supporting datasets and CLM-CNP, *CESM Land Model and Biogeochemistry Working Groups Meeting*, NCAR, Boulder, CO, February 20 – 22, 2013.

Jain, A., B. El-Masri, R. Barman (2013): Carbon exchange in the northern high latitude terrestrial ecosystems over the last three decades, *European Geosciences Union General Assembly 2013*, Vienna, Austria, April 7-12, 2013.

Jain, A.K., B. El-Masri, Y. Jin (2013): Implementation of Prognostic LAI in a land surface model to improve water, energy, and carbon fluxes. US-International Association for Landscape Ecology, Austin, TX, April 14-18, 2013.

ElMasri, B, R Barman, AK Jain (2013), The interaction between biogeophysical and biogeochemical processes and their feedback on permafrost soil carbon stocks, *American Geophysical Union (AGU) Fall meeting*, San Francisco, CA, December 9 – 13, 2013.

Jain AK, Y Jin, B El-Masri, X Yang (2013), Modeling of N2O emissions from soils in terrestrial ecosystems, *American Geophysical Union (AGU) Fall meeting*, San Francisco, CA, December 9 – 13, 2013.

Walker AP, S Zaehle, MG De Kauwe, BE Medlyn, M Dietze, T Hickler, CM Iversen, AK Jain, Y Luo, HR McCarthy, WJ Parton, C Prentice, PE Thornton, S Wang, Y Wang, D Warlind, J Warren, E Weng, PJ Hanson, R Oren, RJ Norby (2013), Model-experiment synthesis at two FACE sites in the southeastern US. Forest ecosystem responses to elevated CO[2], *American Geophysical Union (AGU) Fall meeting*, San Francisco, CA, December 9 – 13, 2013.

Harper A, et al. (Co-authors list include AK Jain) (2014): The carbon cycle response of the Amazon forest during the 2010 drought in dynamic global vegetation models, *Fall 2014 American Geophysical Union (AGU) Meeting*, San Francisco, CA, Dec. 15-19, 2014.

Huntzinger D et al. (Co-Authors list include AK Jain) (2014): Trends in the Global Net Land Sink and Their Sensitivity to Environmental Forcing Factors: Results From the Multi-Scale Synthesis and Terrestrial Model Intercomparison Project (MsTMIP) , *Fall 2014 American Geophysical Union (AGU) Meeting*, San Francisco, CA, Dec. 15-19, 2014.

Schwalm C et al. (Co-authors list include AK Jain) (2014): Attributing Changes in Gross Primary Productivity from 1901 to 2010, *Fall 2014 American Geophysical Union (AGU) Meeting*, San Francisco, CA, Dec. 15-19, 2014.

Shu S, Forrest Hoffman, J Kumar, W Hargrove, A Jain (2014): Data Mining Approach for Evaluating Vegetation Dynamics in Earth System Models (ESMs) Using Satellite Remote Sensing Products, Fall 2014 American Geophysical Union (AGU) Meeting, San Francisco, CA, Dec. 15-19, 2014.

Jain AK, B El-Masri and S Shu (2015): Implementation of Dynamic Rooting Depth and Phenology into a Land Surface Model: Evaluation of Carbon, Water, and Energy fluxes for the High Latitude Ecosystems, *2015 AGU Joint Assembly*, Montreal, Canada, May 4-7 2015.

Ahlström A, MR Raupach, G Schurgers, B Smith, A Arneth, M Jung, M Reichstein, J Canadell, Pierre Friedlingstein, AK Jain, E Kato, B Poulter, S Sitch, B David Stocker, N Viovy, Y Wang, A Wiltshire, S Zaehle, N Zeng (2015). The dominant role of semi-arid lands in the trend and variability of the land CO2 sink, *Fall 2015 American Geophysical Union (AGU) Meeting*, San Francisco, CA, Dec. 13-18, 2015.

Huntzinger DN, AM Michalak, C Schwalm, P Ciais, KM Schaefer, AW King, Y Wei, RB Cook, JB Fisher, DJ Hayes, M Huang, A Ito, AK Jain, H Lei, C Lu, F Maignan, J Mao, N Parazoo, S Peng, B Poulter, DM Ricciuto, X Shi, H Tian, W Wang, N Zeng, F Zhao (2015). Nitrogen Dynamics are a Key Factor in Explaining Global Land Carbon Sink, *Fall 2015 American Geophysical Union (AGU) Meeting*, San Francisco, CA, Dec. 13-18, 2015.

Mao J, W Fu, X Shi, DM Ricciuto, JB Fisher, RE Dickinson, Y Wei, W Shem, S Piao, K Wang, CR Schwalm, H Tian, M Mu, MA Arain, P Ciais, RB Cook, YJ Dai, DJ Hayes, FM. Hoffman, M Huang, S Huang, DN Huntzinger, A Ito, AK Jain, AW King, H Lei, C Lu, AM Michalak, N Parazoo, C Peng, S Peng, B Poulter, KM Schaefer, EE Jafarov, PE Thornton, WWang, N Zeng, Z Zeng, F Zhao, Q Zhu, Z Zhu (2015). Disentangling climatic and anthropogenic

controls on global terrestrial evapotranspiration trends, *Fall 2015 American Geophysical Union (AGU) Meeting*, San Francisco, CA, Dec. 13-18, 2015.

Schwalm C, W Anderegg, F Biondi, GW Koch, ME Litvak, J Shaw, A Wolf, DN Huntzinger, AM Michalak, KM Schaefer, JB Fisher, RB Cook, Y Wei, Y Fang, DJ Hayes, M Huang, AK Jain, H Tian (2015). Global Patterns of Drought Recovery, *Fall 2015 American Geophysical Union (AGU) Meeting*, San Francisco, CA, Dec. 13-18, 2015.

Shu S, U Mishra, FM Hoffman, CD Koven, AK Jain (2015). Interactions Between Soil Organic Carbon Concentration and Soil Thermal and Hydraulic Dynamics and Its Impact on Soil Carbon Storage in Northern High-latitudes, *Fall 2015 American Geophysical Union (AGU) Meeting*, San Francisco, CA, Dec. 13-18, 2015.

Shu, S, F Hoffman, J Kumar, W Hargrove, AK Jain (2016). Synthesis of satellite NDVI products and vegetation phenology in Earth system models using a data mining approach, *US-International Association for Landscape Ecology (IALE) 2016 Annual Meeting*, Asheville, North Carolina, April 3 – 7, 2016.

Thomas RT, IC Prentice, H Graven, P Ciais, JB Fisher, M Huang, DN Huntzinger, A Ito, A Jacobson, A Jain, J Mao, A Michalak, S Peng, B Poulter, DM Ricciuto, X Shi, C Schwalm, H Tian, N Zeng (2016), CO2 and greening observations indicate increasing light-use efficiency in northern terrestrial ecosystems, *European Geophysical Union General Assembly 2016*, Vienna, Austria, April 17–22, 2016.