Modeling spatial–temporal dynamics of global wetlands: comprehensive evaluation of a new sub-grid TOPMODEL parameterization and uncertainties

Z. Zhang¹,², N. E. Zimmermann¹, and B. Poulter²

¹Swiss Federal Research Institute WSL, Dynamics Macroecology, Zürcherstrasse 111, Birmensdorf 8903, Switzerland
²Institute on Ecosystems and Department of Ecology, Montana State University, Bozeman, MT 59717, USA

Received: 17 August 2015 – Accepted: 24 October 2015 – Published: 10 November 2015

Correspondence to: Z. Zhang (yuisheng@gmail.com) and B. Poulter (benjamin.poulter@montana.edu)

Published by Copernicus Publications on behalf of the European Geosciences Union.
Abstract

Simulations of the spatial–temporal dynamics of wetlands are key to understanding the role of wetland biogeochemistry under past and future climate variability. Hydrologic inundation models, such as TOPMODEL, are based on a fundamental parameter known as the compound topographic index (CTI) and provide a computationally cost-efficient approach to simulate wetland dynamics at global scales. However, there remains large discrepancy in the implementations of TOPMODEL in land-surface models (LSMs) and thus their performance against observations. This study describes new improvements to TOPMODEL implementation and estimates of global wetland dynamics using the LPJ-wsl dynamic global vegetation model (DGVM), and quantifies uncertainties by comparing three digital elevation model products (HYDRO1k, GMTED, and HydroSHEDS) at different spatial resolution and accuracy on simulated inundation dynamics. In addition, we found that calibrating TOPMODEL with a benchmark wetland dataset can help to successfully delineate the seasonal and interannual variations of wetlands, as well as improve the spatial distribution of wetlands to be consistent with inventories. The HydroSHEDS DEM, using a river-basin scheme for aggregating the CTI, shows best accuracy for capturing the spatio-temporal dynamics of wetlands among the three DEM products. The estimate of global wetland potential/maximum is \( \sim 10.3 \text{ Mkm}^2 (10^6 \text{ km}^2) \), with a mean annual maximum of \( \sim 5.17 \text{ Mkm}^2 \) for 1980–2010. This study demonstrates the feasibility to capture spatial heterogeneity of inundation and to estimate seasonal and interannual variations in wetland by coupling a hydrological module in LSMs with appropriate benchmark datasets. It additionally highlights the importance of an adequate investigation of topographic indices for simulating global wetlands and shows the opportunity to converge wetland estimates across LSMs by identifying the uncertainty associated with existing wetland products.
1 Introduction

For their ability to emit the greenhouse gas CH$_4$, wetland ecosystems play a disproportionately important role in affecting the global climate system through biogeochemical feedbacks (Fisher et al., 2011; Seneviratne et al., 2010). Wetlands are thought to be the largest natural source of methane (CH$_4$) emission to the atmosphere by contributing 20–40% of the total annual emissions to atmosphere, which adds a strong radiative forcing from CH$_4$ (Bousquet et al., 2006; IPCC, 2013). The seasonal and interannual distribution of wetland area remains one of the largest uncertainties in the global CH$_4$ budget (Kirschke et al., 2013), in particular for the roughly 60% of wetlands that are not inundated permanently (Petrescu et al., 2010). The interannual changes in the distribution of wetlands were most likely a major driver for CH$_4$ variations during last glacial period (Kaplan, 2002) and are considered as an important driver of the strong atmospheric CH$_4$ growth rate resumed in 2007 (Nisbet et al., 2014) and in future climate change scenarios (Stocker et al., 2013).

Improving our understanding of the role of wetlands in global greenhouse-gas (GHG) budgets requires a representation of wetlands and their biogeochemical processes in land surface models (LSM) to both hindcast observed past variations (Singarayer et al., 2011) and to predict future trajectories in atmospheric CH$_4$ and terrestrial C balance (Ito and Inatomi, 2012; Meng et al., 2012; Spahni et al., 2011; Stocker et al., 2014; Zürcher et al., 2013). Dynamic wetland schemes in LSMSs were initially developed from approaches that simulated the upslope contributing area for runoff in hydrologic watersheds. These approaches were based on conceptual theories and physical processes describing surface water processes (e.g., infiltration and evapotranspiration) and water movement in the soil column using probability distributions derived from subgrid topographic information (Beven and Kirkby, 1979), or using analytical functional parametric forms with fixed parameters (Liang et al., 1994). Currently, the most common approach for global wetland modelling is to use a runoff simulation scheme such as TOPMODEL (TOPography-based hydrological MODEL) (Beven and Kirkby, 1979; Kleinen
et al., 2012; Ringeval et al., 2012; Zhu et al., 2014), which includes the assumption that lateral soil water transport as being driven by topography follows an exponential decline of saturated hydraulic conductivities within soil profiles in a basin (Sivapalan et al., 1987).

TOPMODEL-based implementations have proven successful at capturing the broad geographic distribution of wetlands and their seasonal variability (Gedney and Cox, 2003; Ringeval et al., 2012; Stocker et al., 2014; Zhu et al., 2014), but have consistently overestimated the extent and duration of wetlands, both at the global and regional scale when compared with existing current surveys (Junk et al., 2011; Prigent et al., 2007; Quiquet et al., 2015). For instance, simulations using the Earth system model HadGEM2 predict much larger persistent Amazonian wetlands than inventory (Collins et al., 2011). In general, independently determined wetland area using hydrologic modules of LSMs in The Wetland and Wetland CH₄ Inter-comparison of Models Projects (WETCHIMP) experiment simulated larger global wetland extent than those informed by remotely sensed product and inventories (Melton et al., 2013). This large disagreement also exists across specific regions (Ringeval et al., 2014). For example, Bohn et al. (2015a) carried out a model inter-comparison of wetland extent on the West Siberian Lowland, one of the major wetland regions in high latitudes, and highlighted similar uncertainties of wetland extent simulation in the LSMs participating in the WETCHIMP experiment and using TOPMODEL.

Meanwhile, uncertainties in wetland area estimation partly come from a paucity of observational datasets and different definitions of wetland (Matthews and Fung, 1987). Remotely sensed datasets have difficulties in capturing small or isolated water in saturated soils that are not flooded on the surface (Prigent et al., 2007), as well capturing the forested wetlands that obscure detection of inundation because of dense forest canopies (Bohn et al., 2015a). In addition, ground-based survey or inventories that determine wetlands usually limited as static distribution that cannot provide temporal patterns for inundated area, making it hard to evaluate with simulated results. On the other hand, the definition of wetland for regional- or global-scale modelling assumes
the land surface has both inundated and saturated conditions, which is not necessarily
the same as inundated area measured by satellite observations (Melton et al., 2013).

While prognostic wetland dynamics schemes are promising to resolve these observ-

ational issues, the configuration parameters for TOPMODEL are a potential source of
uncertainty in estimating wetland dynamics (Marthews et al., 2015). LSMs are usually
run at coarse spatial resolution (e.g. 0.5°) and the physics they follow is based predom-
nantly on approximations required to scale processes that occur at much finer spatial
resolution (e.g. 10–100 m) to a coarser grid (Ducharne, 2009; Mulligan and Wainwright,
2013). The well-known Compound Topographic Index (CTI), which is widely used in hy-
drology and terrain-related applications (Ward and Robinson), is the key basis describ-
ing topographic information in TOPMODEL. Currently, most of the global applications
derive a CTI product at 1 km resolution from HYDRO1k global dataset released by
USGS in 2000 (Kleinen et al., 2012; Lei et al., 2014; Wania et al., 2013), which has
been proven to at least partly cause biases due to limited spatial resolution (Ringeval
et al., 2012) and also because of the quality of the underlying digital elevation model
(Marthews et al., 2015). These uncertainties will correspondingly lead to inaccurate
estimation in maximum soil water content, as well as in the maximum inundated area
in TOPMODEL.

The primary goal of our study is to improve the modeling of dynamically varying
wetland extents with (i) a parameter constraint to match integrated satellite and inven-
tory observations, and with (ii) a better parameterizations of CTI values for determining
wetland seasonal cycles using new topographic data and aggregation schemes (i.e.,
grid vs. catchment). To this end, we develop a new version of LPJ-wsl that includes the
TOPMODEL approach for wetland extent modelling by also accounting for soil thermal
dynamics and high-latitude soil-water freeze and thaw cycles, and by incorporating
the necessary physical processes that constrain global wetland dynamics. We utilize
three commonly used global DEM products to evaluate the effects of sub-grid parameter-
izations on simulated global wetland extent uncertainties. We perform six global
simulations resulting from the combination of three DEM products and two aggregation
schemes under the same common experimental protocol. The specific aims are: (1) to improve the performance of estimated wetland extent based on TOPMODEL for the purpose of large-scale modelling, (2) to develop a new parameterization scheme using inventory in combination with satellite-based retrievals, and (3) to evaluate the uncertainties and the spatial and temporal differences of CTI from three major DEM products in model behavior.

2 Model descriptions and experimental design

The model LPJ-wsl is a process-based dynamic global vegetation model developed for carbon cycle applications based on development of the LPJ-DGVM (Sitch et al., 2003). LPJ-wsl includes land surface processes, such as water, carbon fluxes, and vegetation dynamics that are intimately represented by plant functional types (PFTs) (Poulter et al., 2011). The distribution of PFTs is simulated based on a set of bioclimatic limits and by plant-specific parameters that govern the competition for resources. The soil hydrology is modeled using semi-empirical approach, with the soil treated as bucket consisting of two layers each with fixed thickness (Gerten et al., 2004). The LPJ-wsl CH4 model used in this study is the same as presented in (Hodson et al., 2011; Wania et al., 2013) as a function of two scaling factors ($r_{CH4:C}$ and $f_{ecosys}$), soil temperature, soil-moisture-dependent fraction of heterotrophic respiration, and wetland extent according to the following equation:

$$
E(x,t) = r_{CH4:C} \cdot f_{ecosys}(x) \cdot A(x,t) \cdot R_h(x,t),
$$

where $E(x,t)$ is wetland CH4 flux, $A(x,t)$ is wetland extent, $R_h(x,t)$ is heterotrophic respiration.

LPJ-wsl has been evaluated in previous studies using global inventory datasets and satellite observations and has been contributed as one of the participating models in the WETCHIMP study (Melton et al., 2013). Modifications to the original LPJ-wsl model and a detailed description of changes are summarized below:
– A permafrost module that simulate soil freeze and thaw processes, is implemented and modified following the Wania et al. (2009) study (see description in Sect. 2.1).

– The snow module from Wania et al. (2009) was included and modified to include some of the effects of snow ageing on snow thermal properties. We use an updated parameterization of soil thermal properties both for the permafrost and the snow module, which is calibrated by satellite observations specifically for global application.

– A new parameterization of soil texture was formulated based on the Harmonized World Soil Database (HWSD), which combines the recently collected extensive volumes of regional and national updates of soil parameter information (Nachtergaele et al., 2008). The new soil texture in LPJ-wsl follows the USDA soil classification with 14 soil types grouped according to a particular range of particle-size fractions (e.g. sand, clay, loam, etc.), instead of using the original FAO classification with 9 soil types (Sitch et al., 2003). Thus, the volumetric water holding capacity, also defined as potential maximum soil water content (SWC), is assumed to vary spatially, calculated as a function of the surface soil texture using pedo-transfer functions from (Cosby et al., 1984). Wilting point, porosity, mineral soil content and organic soil content for each soil type are derived from a look-up table available from the AGRMET (2002) as listed in Table 1.

The modified LPJ-wsl version is thus the starting point upon which the TOPMODEL-based wetland and permafrost modules are included (Sect. 2.2).

2.1 Permafrost model

In order to consider the functional wetland area extension during the spring thaw and their shrinking or disappearances during autumn freeze, we added to LPJ-wsl a soil temperature scheme and freeze–thaw processes, as in Wania et al. (2009). The modi-
fied version considers the soil heat capacity and its thermal conductivity, which are both affected by the volumetric fractions of the soil physical components, such as water-ice fraction, mineral soil, or peat. The thermal scheme of LPJ-wsl is discretized vertically using 8-layers of variable thickness, while the water-balance scheme is kept the same as the original LPJ-DGVM, which means the daily changes in water content are allocated to the “old” upper and lower layer of LPJ while considering percolation between these two layers and baseflow from the lower layer. Fractional water and ice content in each of the 8-layers is calculated on a daily time step. Soil temperature is updated in the thermal routine and then passed to the hydrological routine to determine the water-ice phase change in permafrost routine.

2.2 Dynamic wetland model

To represent the grid cell fraction covered by wetlands, we have implemented an approach based on the TOPMODEL hydrological framework (Beven and Kirkby, 1979). TOPMODEL was initially developed to operate at the scale of large watersheds using the channel network topography and dynamics contributing areas for runoff generation, and was later extended to perform over areas that are much larger than a typical river catchment (Gedney and Cox, 2003). The fundamental information to determine the area fraction with soil water saturation is derived from knowledge of the mean watershed water table depth and a probability density function (PDF) of combined topographic and soil properties (Sivapalan et al., 1987). The Compound Topographic Index, which provides the sub-grid scale topographic information in TOPMODEL, determines the likelihood of a grid box to be inundated. It is defined as:

$$\lambda_l = \ln \left( \frac{a_l}{\tan \beta_l} \right), \quad (2)$$

where $\lambda_l$ represents local CTI value, $a_l$ represent the contributing area per unit contour, $\tan \beta_l$, the local topographic slope, approximates the local hydraulic gradient where $\beta$ is the local surface slope. The CTI distribution can be generated from digital elevation
models and near global datasets are readily available, e.g. HYDRO1k dataset from USGS.

Following the central equations of TOPMODEL, the relationship between local water table depth $z_l$ and the grid mean water table depth $z_m$ can be given as:

$$\lambda_l - \lambda_m = f (z_l - z_m), \quad (3)$$

where $\lambda_m$ is the mean CTI averaged over the grid box, $f$ is the saturated hydraulic conductivity decay factor with depth for each soil type. This equation is valuable in that it relates the local moisture status to the grid box mean moisture status based on the subgrid-scale variations in topography. Higher CTI values than average are indicative of areas with higher water table depth than average water table, and vice versa. We therefore calculate the inundated areas ($F_{\text{wet}}$) of all the sub-grid points within a grid cell that have a local water table depth $z_l \geq 0$:

$$F_{\text{wet}} = \frac{z_{\text{max}}}{z_l} \int_{z_l}^{z_{\text{max}}} pdf(\lambda) d\lambda, \quad (4)$$

where furthermore, instead of using the CTI values themselves, we followed a common up-scaling approach to approximate the distribution of CTI values within a grid cell in order to reduce computation costs. Here, the discrete distribution of the CTI for lowland pixels (i.e. $\lambda_l \geq \lambda_m$) has been represented as an exponential function, not as a three-parameter gamma distribution as applied in recent applications for modeling wetland extent (Kleinen et al., 2012; Ringeval et al., 2012). As shown in Fig. 1, the new exponential function agrees very well with the three-parameter gamma distribution function when the CTI is larger than the mean CTI $\lambda_m$. This change allows linking the inundated fraction directly to water table depth, thus improving the parameterization by providing physical meaning and fewer calibration parameters. This change also improves the parameterization of fractional saturated area, especially in mountainous regions (Niu et al., 2005).
Finally, the wetland area fraction \((s)\) is represented as:

\[
F_{\text{wet}} = F_{\text{wet}}^{\text{max}} e^{-Cs \left(\lambda_l - \lambda_m\right)},
\]

(5)

where \(C_s\) is a coefficient representing the topographic information by fitting the exponential function to the discrete cumulative distribution function (CDF) of the CTI. \(F_{\text{max}}^{\text{wet}}\) is the observed maximum wetland fraction of a grid cell. Because of the uncertainties involved in determining the water table depth, the hydraulic factor \(f\), and the coarse resolution DEMs, the maximum soil saturated fraction calculated from discrete CDF are prone to large uncertainties and thus complicate the comparison of the saturated fraction with existing observations (Ducharne, 2009; Ringeval et al., 2012). Here, we introduce a parameterization in order to calibrate maximum wetland fractions \(F_{\text{max}}^{\text{wet}}\) from fractions at the original maximum saturated fraction \(F_{\text{max}}\), which is calculated from the CDF of CTI when \(\lambda_m\) equals zero. This parameterization is also based on the assumption that water is stagnant within local grids at large scale, in particular for model using simple “bucket” concept to calculate grid-mean water table depth. We used the inventory-calibrated satellite observations SWAMPS-GLWD (see description in 3.3) combining with GLWD to calculate representative long-term maximum wetland extents within each grid box (0.5°), i.e. the parameter \(F_{\text{max}}\) for each grid cell \(i\):

\[
F_{\text{max},i}^{\text{wet}} = \max(A_{\text{GLWD},i}, \max(A_{\text{SWAMP-GLWD},i})).
\]

(6)

\(A_{\text{GLWD},i}\) represents wetland estimate from GLWD, and \(A_{\text{SWAMP-GLWD},i}\) represents long-term wetland estimate from SWAMPS-GLWD. The reason for combining these two datasets is to take the advantage of satellite-based observations at capturing temporal wetlands and inventory-based datasets at estimating forested wetlands and small wetlands ignored by remote sensing.

In addition, we used nonlinear least squares (nls) estimates to fit the CDF curve of CTI only for lowlands \((\lambda_l < \lambda_m)\) to calculate parameter \(C_s\), the parameter that determines varying trend of wetland extent. By this, the parameters \(F_{\text{max}}, \lambda_m\) and \(C_s\) for determining inundated areas are derived (Fig. 2).
To account for the permafrost effects on soil infiltration properties, we followed Fan and Miguez-Macho (2011) and Kleinen et al. (2012) who modified \( f \) by a function \( k \) depending on January temperature \( T_{\text{jan}} \). Since LPJ-wsl uses two soil layers from the HWSD soil texture database to represent the different texture characteristics, the modification depends on the combination of a look-up table (Table 1) from soil types and water table depth:

\[
k = \begin{cases} 
1 & \forall T_{\text{jan}} > -5 \\
1.075 + 0.015T_{\text{jan}} & -25^\circ < \forall T_{\text{jan}} < -5^\circ \text{C} \\
0.75 & \forall T_{\text{jan}} < -5^\circ \text{C}
\end{cases}
\]  

(7)

Since the observed CH\(_4\) emission during winter are more attributed to physical processes during soil freezing effects (Whalen and Reeburgh, 1992), for the partially frozen wetland in high latitude, we introduced an effective fraction of wetland area \( (F_{\text{wet}}^{\text{eff}}) \) defined by:

\[
F_{\text{wet}}^{\text{eff}} = \left( \frac{\omega_{\text{liq}}}{\omega_{\text{liq}} + \omega_{\text{froz}}} \right)_{50 \text{ cm}} \cdot F_{\text{wet}},
\]

(8)

where \( \omega_{\text{liq}} \) and \( \omega_{\text{froz}} \) are the fraction of liquid and frozen soil water content in the upper soil (0–0.5 m) respectively. Since the liquid water content in the lower soil layer gets trapped and cannot contribute to CH\(_4\) emission when upper soil is frozen, we did not consider the lower layer for surface wetland calculations.

3 Experimental set-up and datasets

3.1 Topographic information

In this study we used three DEMs of varying spatial resolution, HYDRO1k (USGS, 2000; https://lta.cr.usgs.gov/HYDRO1K), Global Multi-resolution Terrain Elevation Data
2010 (GMTED) (Danielson and Gesch, 2011), and HydroSHEDS (Lehner et al., 2008) to compare the effect of sub-grid topographic attributes on simulated seasonal and interannual variability of wetlands. HYDRO1k, developed from the USGS released 30 arcsec digital elevation model of the world (GTOPO30), is the first product that allowed spatially explicit hydrological routines applied in large-scale applications (USGS, 2000). HydroSHEDS, developed from satellite-based global mapping by the Shuttle Radar Topography Mission (SRTM), is a significant improvement in the availability of high-resolution DEMs covering all land areas south of 60° N (the limit of SRTM). For the areas at higher latitudes we used HYDRO1k by aggregating the GTOPO30 DEM to provide global grids. GMTED was produced using seven data sources including SRTM, global Digital Terrain Elevation Data (DTED), Canadian elevation data, Spot 5 Reference3-D data, and data from the Ice, Cloud, and land Elevation Satellite (ICESat), covering nearly all global terrain.

In order to account for uncertainties inherent in computing CTI with different CTI algorithms, we generated a global CTI map based on the three DEM products, instead of relying on existing CTI products. Since studies show that multiple flow direction algorithms for calculating CTI give better accuracy compared with single-flow algorithms in flat areas (Kopecký and Čížková, 2010; Pan et al., 2004), thus we selected an algorithm from R library “topmodel” (Buytaert, 2011), which applies the multiple flow routing algorithm of Quinn et al. (1995) to calculate the global CTI maps. The DEMs from HYDRO1k and HydroSHEDS had been previously processed for hydrological-correction, meaning that the DEMs were processed to remove elevation depressions that would cause local hydrologic “sinks”. To include a comparison of (hydrologically) corrected and uncorrected DEMs in our analyses as some studies have been done previously (Stocker et al., 2014), we retained the GMTED DEM without hydrologically correction.

One of key assumptions in TOPMODEL is that the water table is recharged at a spatially uniform and steady rate with respect to the flow response timescale of the catchment (Stieglitz et al., 1997). Given the fact that we consider the water to be stagnant within each grid, the mean CTI parameter was estimated with two alternative schemes:
(1) a regular “tile-based” or gridded approach, i.e., the subgrid CTI values were averaged per 0.5° tiles, and (2) an irregular “basin-based” approach, where mean CTI were calculated over the entire catchment area in which the respective pixel is located. For generating global catchment map at 0.5° resolution, we applied a majority algorithm in the case of multi-catchments in a tile with consideration of avoiding isolated pixels for specific river basin. There are two catchment area products applied in this study, HYDRO1k (2013) and HydroSHEDS. Similarly, the parameter $C_s$ was generated using nonlinear least squares estimates from both of these two different CTI calculation strategies. The descriptions of DEM products are summarized in Table 2.

3.2 Description of the simulation

For running LPJ-wsl with permafrost and TOPMODEL, we used global meteorological forcing (temperature, cloud cover, precipitation and wet days) as provided by the Climatic Research Unit (CRU TS 3.22) at 0.5° resolution (Harris et al., 2014). To spin up the LPJ-wsl model using the CRU climatology, climate data for 12 months were randomly selected from 1901–1930 and repeated for 1000 years with a fixed pre-industrial atmospheric CO$_2$ concentration. The first spinup simulation started from initial soil temperature derived from LPJ-wsl simulated results on January 1901 and continued with a land use spinup simulation. These procedures ensure that carbon stocks and permafrost are in equilibrium before performing transient simulations. The transient simulations, with observed climate and CO$_2$ were performed with monthly climate disaggregated to daily time steps over the 1901–2013 period. Two sets of model experiments were carried out to compare the wetland dynamics under basin and tile-based TOPMODEL parameterizations respectively. The 1993–2013 years were used for evaluation against satellite data and inventories.
3.3 Evaluation and benchmarking data

Since the soil freeze–thaw cycles are a key component for determining seasonal cycles of wetlands in cold regions, in this study we benchmarked the general pattern of permafrost locations by comparing the model output against satellite observations of freeze and thaw status and inventories of permafrost extent. Since soil depth in LPJ-wsl is held at 2.0 m for the permafrost module, the permafrost extent in this study is defined as the lower soil (0.5–2 m) that is always at or below the freezing point of water 0°C for multiple years. The permafrost extent map at 0.5° resolution from National Snow and Ice Data Center (NSIDC) is adopted for benchmarking (Brown et al., 2001). The global dataset of Freeze/Thaw (FT) status is derived from satellite microwave remote sensing provided by the Numerical Terradynamic Simulation Group (NTSG) at University of Montana and is based on daily maps over a 34 year record (1979–2012). It represents the FT status of the composite landscape vegetation-snow-soil medium to constrain surface water mobility and land–atmosphere carbon fluxes (Kim et al., 2012).

Two global inundation products derived from satellite observations were additionally used for evaluation purposes: the Global Inundation Extent from Multi-Satellites (GIEMS), derived from visible (AVHRR) and active (SSM/I) and passive (ERS) microwave sensors over the period 1993–2007; the Surface Water Microwave Product Series (SWAMPS), derived from active (SeaWinds-on-QuikSCAT, ERS, and ASCAT) and passive (SSM/I, SSMI/S, AMSR-E) microwave sensors over the period 1992–2013. This new SWAMPS global dataset, hereby denoted as SWAMPS-GLWD, was first developed at NASA JPL (Schroeder et al., In preparation). We re-scaled this dataset with the Global Lake and Wetland Database (GLWD) (Lehner and Döll, 2004), a well-established global inventory of water bodies at high resolution to match SWAMPS-GLWD with the inventory estimates. This post-processed SWAMPS product covers the required regions for forested wetlands, which are not readily observable by passive or active microwave measurements (Poulter, 2015). For evaluating regional wetland patterns, we selected two study areas (the largest peatland West Siberian Lowland
(WSL); the largest floodplain, Amazon River Basin). Three wetland map products over the WSL from Sheng et al. (2004), Peregon et al. (2008) and Tarnocai et al. (2009) (denoted by “Sheng2004”, “Peregon2008”, “Tarnocai2009” respectively) and one update high resolution dual-season inundated area inventory for lowland Amazon basin from Hess et al. (2015) (denoted by “Hess2015”) were applied. We aggregated all above-mentioned datasets from the native 25 km to a 0.5° spatial resolution and from daily to monthly temporal resolution for comparison with model outputs (Table A1 in the Appendix).

4 Results

4.1 Evaluation against observations

We first evaluated the permafrost module that constrains the seasonal cycles of wetland area in cold regions with respect to inventory and remote sensing observations. Figure 3a compares the spatial distribution of permafrost extent from inventory and the modeled permafrost extent over the period 1980–2000. Figure 3b gives the spatial distribution of spearman rank correlation between the simulated and observed number of monthly frozen-days. The modeled permafrost extent shows high agreement with benchmarking dataset, with a slightly higher coverage of permafrost regions in North-Western Eurasia. The model successfully captures the seasonal frozen soil, which is closely linked to surface wetland formation and seasonal variation of wetland in cold regions. Most of the regions reveal a temporal correlation > 0.9, while Eastern Siberia and the Southern permafrost distribution edge is generally around 0.5. The lower correlation in East Siberia probably originates from two issues: high snowpack in LPJ-wsl that insulates soil temperature and prevents complete freezing; and the relatively large uncertainty of FT-ESDR derived soil frozen status in those region (Kim et al., 2012). This difference can be partly explained by the different representation of frozen status between simulated results and satellite retrievals. Remotely sensed maps reflect the
mixed condition of the upper vegetation canopy, snow layer and surface soil, while the simulated frozen days only represent the frozen state of topsoil.

Figure 4 illustrates the model evaluation at the regional scale over the West Siberian Lowland (Fig. 4). The model generally captures the spatial extent of the seasonal maximum wetland area fraction across the whole WSL for the JJA season successfully. However, the TOPMODEL approach without calibration (denoted as “Original”) shows large areas with relatively low wetland proportion and cannot capture high values. This suggests poor model performance in simulating wetland areas without $F_{\text{max}}$ calibration. The calibrated model generally exhibits good agreement with inventories and satellite retrievals. It is especially successful at capturing the spatial heterogeneity of wetland areal extent over the whole WSL regions. LPJ-wsl simulated results reveal additional wetland area in the northeast, where wetlands entirely lacked in the GLWD map, although captured in other datasets. Meanwhile, LPJ-wsl captured the higher wetland area in region between 61 and 66° N and 70 and 80° E regions compared with GLWD, where mire/bog/fen was dominated across that region. LPJ-wsl also maintained well the spatial pattern of wetlands in forested region south of 60° N, which was captured by inventories (Sheng2004; Peregon2008, and GLWD), but was missed by two satellite products (SWAMPS-GLWD, GiEMS) due to the limitation of remotely sensed datasets in detecting water under vegetative canopy and/or due to reduced sensitivity.

As illustrated in Fig. 5, LPJ-wsl captured the spatial pattern of simulated wetlands well with lower estimates of the total wetland area in low-water season compared to the JERS-1 observed maps. Differences between Hess2015 and LPJ-wsl maps were primarily in two regions, Maranon-Ucayali region of Peru (MUP, 3–7° S, 73–77° W) and Llanos de Moxos in Bolivia (LMB, 11–17° S, 60–68° W). LPJ-wsl shows higher wetland coverage in MUP while Hess2015 indicates high wetland fraction in LMB in high-water season. Global satellite products largely ignore the LMB region that was partly captured in LPJ-wsl, indicating that LPJ-wsl using hybrid TOPMODEL approach can yield estimates closer to those of fine-resolution mapping, while large-scale satellite products are likely to underestimate Amazon wetland extent because of their coarse spatial
resolution that limit the ability to detect inundation outside of large wetlands and river floodplains (Hess et al., 2015).

In order to evaluate the effect of wetland parameterization on CH$_4$ emission estimates, two estimates of CH$_4$ from LPJ-wsl over the WSL regions were compared with observation-based estimate from Glagolev et al. (2011) (Fig. 6). The 3-year mean estimates on annual total emission from non-calibrated TOPMODEL is $6.29 \pm 0.51$ Tg CH$_4$ yr$^{-1}$, falling into the upper part of range from land surface models and inversions (Bohn et al., 2015b), while the calibrated version maintains lower level of CH$_4$ emission with $4.07 \pm 0.45$ Tg CH$_4$ yr$^{-1}$, which is close to the estimate of Glagolev et al. (2011) ($3.91 \pm 1.29$ Tg CH$_4$ yr$^{-1}$). In addition, calibrated TOPMODEL reproduces a good spatial pattern with relatively stronger emissions in Taiga forests and majority of emission in central region (55–65° N, 65–85° E). The non-calibrated result shows relatively less spatial variability in emission, likely due to the area bias of simulated wetlands. We also compared our estimate with recent CARVE airborne observations for Alaska during 2012. Our calibrated TOPMODEL also falls well into the range of recent estimate ($2.1 \pm 0.5$ Tg CH$_4$ yr$^{-1}$) for Alaska based on airborne observations (Chang et al., 2014) with a total of 1.7 Tg CH$_4$ yr$^{-1}$ during 2012 growing season ($3.1$ Tg CH$_4$ yr$^{-1}$ from non-calibrated estimate), indicating the capability of our approach to accurately capture annual CH$_4$ emission and spatial variability for boreal wetlands.

### 4.2 Spatial distribution

Several observations applicable to evaluate the difference among sub-grid parameterizations of TOPMODEL are available for the WSL region. Figure 7 lists the spatial patterns of simulated JJA wetland area over WSL regions to illustrate differences among wetland maps. The general patterns of wetland extent are substantially similar, because they both used the same calibrated $F_{\text{max}}$ map. Both of these datasets show wetlands distributed across most of the WSL, with extensive wetlands in the central region (55–65° N, 60–90° E). However, the detailed pattern is differing between the ap-
proaches and DEMs used, which indicate the uncertainty of parameterizations on wetland distribution. The basin-based parameterization can capture the higher wetland areas in regions with bog, mire, or fen vegetation in the central east (63–67° N, 85–90° E) as was found in the GLWD benchmark map. The tile-based parameterizations fail to reproduce this pattern. It seems that the tile-based parameterizations are less sensitivity in capturing the spatial heterogeneity throughout most of the WSL. The difference in parameterization derived from DEM datasets also affects the simulated regional pattern. Both of HydroSHEDS-based results successfully reproduce the high wetland fractions in the southern-forested regions (55–60° N, 65–80° E), while HYDRO1k and GMTED both cannot capture this feature. Note that GMTED and HydroSHEDS are derived from the same DEM product SRTM for the region lower than 60N, indicating the importance of hydro-correction in simulating spatial patterns of wetlands.

The comparison of simulated mean annual minimum, maximum, and amplitude of wetland extent with observational datasets (Table 3) reveals that the simulated wetland area for 1980–2010 falls within the range of 4.37 ± 0.99 Mkm$^2$ (Mkm$^2 = 10^6$ km$^2$). This number is close to GIEMS (5.66 Mkm$^2$) (Prigent et al., 2012) and inventory-based estimation (6.2 Mkm$^2$) (Bergamaschi et al., 2007) after exclusion of other water bodies like lakes, rivers, and rice paddy (Leff et al., 2004). Considering potential underestimation of satellite-based observation in forested regions, the realistic estimate could possibly be in the upper part of our range. Note that one must be careful when comparing model results with the observational datasets based on inventories or digitalized maps directly, because these datasets might represent the long-term maximal area as wetland potential. The higher seasonal wetland extent in GIEMS compared with LPJ-wsl could be partly due to the artifacts in data retrieval and processing that enlarge the amplitudes in northern high latitudes, and partly due to permanent wetlands that GIEMS hard to detect. Lastly, the definition of wetland is another possible source of discrepancy. Remotely sensed inundation datasets represents the non-specific measurement of inundation while wetland area in our study is specifically defined from inventories.
following the National Wetlands Working Group (1988) classification that include peatlands, mineral wetlands, and shallow waters.

### 4.3 Seasonal cycle

The shapes of the seasonal patterns in wetland area are generally similar in model simulation compared to satellite observations, despite disagreement in the timing of the seasonal cycle of wetland area in some boreal regions (Fig. 8). The modeled results show slightly larger wetland areas in the September–October–November (SON) seasons than satellite-based observations. This is because satellite observations used here does not distinguish between the absence of inundation and masked estimates resulting in some important data gaps in high latitude (Fluet-Chouinard et al., 2015; Papa et al., 2010; Prigent et al., 2012). The higher seasonal wetland areas during SON may originate from the longer unfrozen seasons and relatively saturated soil status in models. It thus seems realistic that the satellite-based inundation product AMSR-E observed a similar trend of seasonal inundation patterns for North America and Boreal Eurasia (Jennifer et al., 2014). This is also supported by field studies in boreal regions, indicating that water table depth during the SON seasons is still in a high level and soil temperature is above freezing status (Rinne et al., 2007; Turetsky et al., 2014).

In contrast, the modeled seasonal cycle of wetland in tropical and temperate regions show a good agreement with GIEMS and SWAMPS-GLWD. Given the difficulties of satellite-based observations in detecting wetlands in forested regions and the reduced sensitivity where open water fraction is low (< 10 %) (Prigent et al., 2007), the inundation numbers by GIEMS might slightly underestimated the area compared with the simulated results.

Figure 8 reveals that the six estimates (base on different DEM processing approaches) of monthly wetland extent averaged for 1993–2007 show the same general behavior in the different regions. The six data sets are highly correlated, with largest differences at the maximal wetland extents during growing seasons, especially in the boreal regions. In addition, the differences in seasonal cycle among the six model ex-
experiments are relatively small, mostly below 5% regardless of the month. This indicates that the averaged total wetland area is not dependent on the introduction of the new sub-grid parameterizations at the global scale. Among the DEM datasets, HYDRO1k shows the largest difference between basin and tile-based estimates with annual mean wetland area of 89,663 km² in boreal regions, while HydroSHEDS has a lowest difference of 6550 km² between the two versions. Examining the seasonal amplitude for basin-based schemes, HydroSHEDS shows a better agreement with satellite-based observations than the other two datasets.

4.4 Interannual variability

For evaluating the performance of all the sub-grid parameterizations, we calculated Pearson’s correlation coefficient ($r$) between modeled and satellite-based results (Table 4). Generally, the comparison demonstrates that simulated interannual variability shows a good agreement with GIEMS and SWAMPS-GLWD in most of the Transcom regions. For boreal and tropical regions, all correlation coefficients are ranging from 0.7–0.8. The comparison of the inter-annual trends (Fig. A1 in the Appendix) indicates that absolute values of simulated interannual variations are close to satellite-based observation with good agreement in shape and timing in these regions. This demonstrates the ability of TOPMODEL to capture the large-scale variations in wetland/inundation. Highest disagreements are found in temperate regions that are strongly affected by human activities (likely strong global anthropogenic effect on continental surface freshwater), which is indicated by GIEMS (Prigent et al., 2012) but not by modeled results.

The interannual variability originating from six different sub-grid DEM parameterizations is very similar between these schemes with Spearman rank correlation coefficient $r > 90\%$. Among the six schemes, the parameters calculated from HydroSHEDS using basin-based statistics result in better agreement between simulated and measured wetland area than the other schemes. In most regions, the SWAMPS-GLWD and GIEMS are consistent in their observed wetland area patterns, except for temperate regions (e.g. Temperate South America, Temperate North America, Europe).
This confirms that the differences in surface water extent detection between GIEMS and SWAMPS-GLWD, which might be caused by observational behaviors from different satellite instruments and algorithms. In addition, parameters estimation based on river basins are slightly better than tile-based results.

5 Discussion

5.1 Wetland modelling based on TOPMODEL concept

The coupling between LPJ-wsl and TOPMODEL with parameter calibrations as described in this study allows for simulating the wetland dynamics, as well as its specific location and extent. The improvement in this study that importing $F_{\text{max}}$ calibration using inventories is based on the recent discussions of the suitability of TOPMODEL application to simulate wetland variations at large spatial scale (Ringeval et al., 2012), and intercomparisons of the wetland-area-driven model bias in CH$_4$ emission at regional scale (Bohn et al., 2015a). The naturally inundated areas simulated by TOPMODEL so far have shown extensive disagreement with inventories and remotely sensed inundation datasets (Melton et al., 2013), and are said to be difficult to validate in absolute numbers. Moreover, these large discrepancies of wetland areas among LSMs were observed, partly due to large varieties of schemes used for representing hydrological processes, and partly due to the inappropriate parameterizations for simulating inundations. To solve this challenges at the global scale, we presented an improved representation of wetland/inundation in LSMs that can be make comparable with benchmark dataset in absolute values is necessary for global wetland modelling.

The simulation of hydrological dynamics within LSMs remains relatively simple because the physics they follow is based predominantly on approximations of processes that occur at much finer spatial scales (Ducharne, 2009; Mulligan and Wainwright, 2013). The coupling of TOPMODEL with process-based LSMs allows for retrieving the fraction at maximum saturated fraction ($F_{\text{max}}$), which is defined by the pixels with no
water deficit estimated from the partial integration of the spatial distribution of CTI in a catchment. This estimated distribution of $F_{\text{max}}$ is much larger than that obtained from the satellite-based observations (Papa et al., 2010). As a key parameter for determining the soil saturated area, the calculation of $F_{\text{max}}$ at large scale is prone to large uncertainties, in particular linked to uncertainties in topographic information, as well as the hydrological processes implemented in large-scale LSMs. Ringeval et al. (2012) pointed to the difficulty of two-layer bucket hydrological model in estimating the mean deficit to the saturation over each grid-cell. This can lead to nonrealistic absolute values of the contributing area in a watershed. We constructed several strategies for optimizing $F_{\text{max}}$ by correcting topographic information to match the wetland inventories (Gedney and Cox, 2003; Kleinen et al., 2012). This is one possible solution for global wetland modeling as it assumes that wetland area can be considered constant at coarse spatial resolution (e.g. 0.5° or 1°), following the classical approach of Beven and Kirkby (1979). However, due to the uncertainties from topographic information used in global application and due to limitations in model parameterization, this approximation cannot capture the fine-scaled wetland extent, which makes comparisons with inventories difficult.

Another different solution adopted in this study is to improve the saturated areas in TOPMODEL by introducing inventory-based dataset to provide regional constraints to wetland area modelling. This modification relies on that, in global application, the TOPMODEL approach does not involved in representation of the global water budget in hydrological module of LSMs (e.g. two-layer bucket model) realistically without such constraining information meant to reduce uncertainties. This hybrid approach for the wetland modelling allows for detecting of “intermittent wetlands” in arid or semiarid regions due to extreme precipitation events and is capable of simulating the wetland dynamics on decade-to-century long time scales. As shown in Fig. 9, the wetland potential for permafrost and arid/semi-arid regions is high. Even in tropical regions, there is ca. 20–30% of potential for terrain to be inundated. Based on the approach that the wetland potential based solely on topographic information within grid cells (meaning that the mean grid cell water table depth is zero), our estimation of global wetland po-
tential/maximum is $\sim 10.3 \text{Mkm}^2$, which comes very close to the deduction ($10.4 \text{Mkm}^2$) from recent estimates at finer resolution for total open water ($\sim 17.3 \text{Mkm}^2$) (Fluet-Chouinard et al., 2015), lakes ($\sim 5 \text{Mkm}^2$) (Verpoorter et al., 2014), and rice paddies ($1.9 \text{Mkm}^2$) (Leff et al., 2004).

According to our evaluation using satellite-based observations and inventories, the spatial distribution of the wetland areas and its temporal variability are generally well captured by our model, both at regional and global scales. In addition, the modeled wetland areas and interannual variability compare well with inventories and satellite-based observations respectively. Unfortunately, the wide disagreement in simulated wetland dynamics among estimates from WETCHIMP hampers our ability to assess model performance (Bohn et al., 2015a). Narrowing down the uncertainty of wetland areas by existing maps could minimize the controversial use of the definition between wetlands and inundations. Wetlands have considerable variations in hydrologic conditions, size, locations that make the wetland definition hard to consistent. In current parameterization, the connectivity of wetlands cannot be represented since wetlands are considered invariant within grid cells.

### 5.2 CTI parameterizations

Among all parameters in TOPMODEL, the compound topographic index (CTI) is of critical importance for determining inundated area in terrain-related hydrological applications (Ward and Robinson, 2000; Wilson and Gallant, 2000). It measures the relative propensity for soils to become saturated (Beven and Cloke, 2012) and consequently it drives the accuracy of wetland area scaled to the larger grid cell. Although the importance of CTI has been highlighted, only few studies have so far evaluated the effect of CTI on modelling the spatial and temporal patterns of global wetland dynamics. This is due to a limited availability of global CTI products and limitation therein. HYDRO1k has become the most commonly applied global dataset for large-scale applications during the last decade. With the development of hydrological routines in LSMs over recent
years, many attempts to improve the accuracy of these fundamental parameters have emerged, from regional or global scales (Grabs et al., 2009; Lin et al., 2010, 2013; Sørensen and Seibert, 2007; Stocker et al., 2014). However, due to the different algorithms applied and study areas in these studies, the CTI distributions are not directly comparable.

As shown in this study, global wetland simulations can benefited from improved spatial resolution of topographic maps, thus creating more realistic representation of processes at sub-grid resolution, and correspondingly better inundation simulations. This is support the ideas of Wood et al. (2011) who claimed that higher-resolution modeling leads to better spatial representation of saturated and nonsaturated areas, even though limitations in up-scaling parameterizations may potentially outrun this advantage. The comparison between HydroSHEDS and GMTED also indicated that, for capturing inundated areas under the same spatial resolution, the parameter maps derived from DEM without hydrological corrections have less accuracy compared to corrected ones (Lehner and Grill, 2013). Without hydrological corrections, valleys would appear as closed depression in the DEM, leading to an underestimation of inundated areas (Marthews et al., 2015). It could be foreseen that if DEMs in process-based models are being applied at higher resolution, this drawback could be amplified. The comparison between basin- and tile-based parameterizations suggests that tile-based calculations are not appropriate and consequently underestimates wetland areas even when assuming invariant inundated areas at large scale.

In addition, the algorithm to calculate CTI is another potential source of error for modelling inundations. The method we applied here is based on calculating a CTI distribution map using a simple algorithm in the R package “topmodel” instead of using an existing CTI product with improved contributing area. The algorithm we applied using the multi-flow direction algorithm that allows for multiple in-flow and out-flow of water among neighboring pixels when generating topographic values. This could potentially overestimate the contributing areas (Pan et al., 2004). As a results, it might underestimate the wetland areas within each grid cell, and slightly underestimate the temporal
pattern of saturated areas because of improper estimates of parameter $C_s$ (Güntner et al., 2004). One limitation of HydroSHEDS is that its projection is not equal-area like HYDRO1k (Marthews et al., 2015), and will cause a potential bias in slope calculation along east–west directions at high latitudes. However, since there is no common method to calculate slope or flow direction, we believe that our calculations provide a reasonable approximation for global applications.

### 5.3 Future needs for global wetland modelling

Substantial progress has been made in the development of wetland modeling, but the wide disagreement among estimates from LSMs still exists (Bohn et al., 2015a; Melton et al., 2013). Considering that spatio-temporal variation of wetland area can largely influence CH$_4$ emissions, the selection of appropriate maps needs to be done with care. The parameterization and evaluation of multi-resolution topographic products presented in this study would enhance global wetland modeling if progress could be made in four areas particularly:

- **Improved parameters of TOPMODEL for large-scale application.** Our results demonstrate that model simulation after calibrating TOPMODEL are comparable in absolute value with inventories and satellite-based observations at coarser resolution. This supports the ideas of (Beven and Cloke, 2012) that an appropriate scale-dependent subgrid parameterization is the main challenge, regardless of whether it is carried out at global modeling scales or landscape scales. The saturated soil water content is the decisive unit that determines wetland distributions and reasonable estimates of global wetland areas. Hydraulic parameters, which describes soil characteristics for water movement, are critical for modelling wetland seasonal cycles (Marthews et al., 2014). Assessing the uncertainties introduced by aggregating sub-pixel to pixel areas also needed to be evaluated.

- **Implementing human impact within wetland modeling.** There are evidences from long-term satellite-based observations detecting a significant effect of human ac-
tivities on wetland drainage at continental scale (Prigent et al., 2012). At finer scale, the variability of wetland extent has also been affected by land-use change (e.g. wetland restoration, deforestation, drainage for forestry, agriculture, or peat mining) and consequently influences spatio-temporal patterns of CH₄ emission (Petrescu et al., 2015; Zona et al., 2009). Land-use change may therefore feedback water available to wetlands through altering water balance between land surface and atmosphere (Woodward et al., 2014). An implementation of human impacts within LSMs at large scale may be important for accurate estimation of interannual variations of wetlands.

- **Improved modelling of soil moisture.** The quality of soil moisture simulation using LSMs depends largely on the accuracy of the meteorological forcing data, surface–atmosphere interaction schemes, and a wide range of parameters (e.g. albedo, minimum stomatal resistance, and soil hydraulic properties). As the fundamental variable for determining water table depth at global scale (Fan et al., 2013), soil moisture plays a key role in simulating the spatio-temporal variability of wetland dynamics. Since it is impossible to produce accurate large-scale estimates of soil moisture from in situ measurement networks (Bindlish et al., 2008; Dorigo et al., 2011), simulation combined with long-term surface and root zone remotely sensed estimates (de Rosnay et al., 2013; Kerr et al., 2010) via data assimilation technology, represents a strategy to improve the capturing of global wetland variability. Future hydrology-oriented satellite missions such as Soil Moisture Active Passive (SMAP) (Entekhabi et al., 2010), and Surface Water and Ocean Topography (SWOT) mission (Durand et al., 2010) are expected to provide soil moisture and will improve the capacity of global soil moisture simulations.

- **Improved satellite benchmark observations.** Current satellite-based estimates of wetland area remain generally uncertain, despite being important for monitoring global wetland variability. Remotely sensed global inundation are prone to underestimate areas of wetlands that are small inundated, as well as covered with
dense vegetation canopies (Papa et al., 2010). Moreover, estimated coastal areas show large bias due to interference with the ocean surface (Prigent et al., 2007). This raises requirement for benchmark dataset to generate more accurate products with lower uncertainties. Downscaling methodology has been made to refine existing satellite-based inundation estimates by coupling the mapping process with reliable inventories (Fluet-Chouinard et al., 2015). This may improve global inundation products, as well as the TOPMODEL parameter estimation in the future.

6 Conclusion

The new LPJ-wsl version incorporates a TOPMODEL approach and a permafrost module representing soil freeze–thaw processes to simulate global wetland dynamics. Once the \( F_{\text{max}} \) parameter in TOPMODEL was calibrated against a benchmark dataset, the model successfully mapped regional spatial pattern of wetlands in West Siberian Lowland and lowland Amazon basin, and captured the spatio-temporal variations of global wetlands well. The parameterization of TOPMODEL based on three DEM products, HYDRO1k, GMTED, and HydroSHEDS revealed that HydroSHEDS performed best in capturing the spatial heterogeneity and interannual variability of inundated areas compared to inventories. River-basin based parameterization schemes using HYDRO1k and GMTED marginally but significantly improve wetland area estimates. The estimates of global wetland potential/maximum is \( \sim 10.3 \text{ Mkm}^2 \), with a mean annual maximum of \( \sim 5.17 \text{ Mkm}^2 \) for 1980–2010. This development of the wetland modeling method reduces the uncertainties in modeling global wetland area and opens up new opportunities for studying the spatio-temporal variability of wetlands in LSMs that are directly comparable with inventories and satellite datasets.

Acknowledgements. Z. Zhang acknowledges funding by the CCES MAIOLICA2 project no. 42-01 and the National Natural Science Foundation of China (Y411391001). We are grateful to J. O. Kaplan and T. J. Bohn for providing WETCHIMP-WSL experiment results. We thank...
B. D. Stocker for providing HYDRO1K global river basin map. We thank T. Mathews for providing global CTI dataset of HydroSHEDS. We thank L. L. Hess for the results of dual-season inundated product of Lowland Amazon Basin.

References


Modeling spatial–temporal dynamics of global wetlands

Z. Zhang et al.


17988


**Table 1.** Soil parameters for LPJ-wsl soil classes. *f* is a parameter describing the exponential decline of transmissivity with depth for each soil type.

<table>
<thead>
<tr>
<th>Soil type</th>
<th><em>f</em></th>
<th>Mineral content (%)</th>
<th>Organic content (%)</th>
<th>Wilting point (%)</th>
<th>Porosity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clay heavy</td>
<td>3.2</td>
<td>0.508</td>
<td>0.01</td>
<td>0.138</td>
<td>0.138</td>
</tr>
<tr>
<td>Silty clay</td>
<td>3.1</td>
<td>0.531</td>
<td>0.01</td>
<td>0.126</td>
<td>0.468</td>
</tr>
<tr>
<td>Clay</td>
<td>2.8</td>
<td>0.531</td>
<td>0.01</td>
<td>0.138</td>
<td>0.468</td>
</tr>
<tr>
<td>Silty clay Loam</td>
<td>2.9</td>
<td>0.534</td>
<td>0.01</td>
<td>0.120</td>
<td>0.464</td>
</tr>
<tr>
<td>Clay loam</td>
<td>2.7</td>
<td>0.595</td>
<td>0.01</td>
<td>0.103</td>
<td>0.465</td>
</tr>
<tr>
<td>Silt</td>
<td>3.4</td>
<td>0.593</td>
<td>0.01</td>
<td>0.084</td>
<td>0.476</td>
</tr>
<tr>
<td>Silt loam</td>
<td>2.6</td>
<td>0.593</td>
<td>0.01</td>
<td>0.084</td>
<td>0.476</td>
</tr>
<tr>
<td>Sandy clay</td>
<td>2.5</td>
<td>0.535</td>
<td>0.01</td>
<td>0.100</td>
<td>0.406</td>
</tr>
<tr>
<td>Loam</td>
<td>2.5</td>
<td>0.535</td>
<td>0.01</td>
<td>0.066</td>
<td>0.439</td>
</tr>
<tr>
<td>Sandy clay Loam</td>
<td>2.4</td>
<td>0.565</td>
<td>0.01</td>
<td>0.067</td>
<td>0.404</td>
</tr>
<tr>
<td>Sandy loam</td>
<td>2.3</td>
<td>0.565</td>
<td>0.01</td>
<td>0.047</td>
<td>0.434</td>
</tr>
<tr>
<td>Loamy sand</td>
<td>2.2</td>
<td>0.578</td>
<td>0.01</td>
<td>0.028</td>
<td>0.421</td>
</tr>
<tr>
<td>Sand</td>
<td>2.1</td>
<td>0.578</td>
<td>0.01</td>
<td>0.010</td>
<td>0.339</td>
</tr>
<tr>
<td>Organic</td>
<td>2.5</td>
<td>0.01</td>
<td>0.20</td>
<td>0.066</td>
<td>0.439</td>
</tr>
</tbody>
</table>
**Table 2.** Model experiments for different parameterization schemes and corresponding DEM products applied in this study.

<table>
<thead>
<tr>
<th>Model experiment</th>
<th>DEM source</th>
<th>Resolution (arcsec)</th>
<th>Coverage</th>
<th>River basin</th>
<th>Aggregation type</th>
<th>Hydro-corrected</th>
</tr>
</thead>
<tbody>
<tr>
<td>HYDRO1k_BASIN</td>
<td>Hydro1k</td>
<td>GTOPO30</td>
<td>30</td>
<td>Global*</td>
<td>HYDRO1K</td>
<td>Catchment</td>
</tr>
<tr>
<td>HYDRO1k_GRID</td>
<td>Hydro1k</td>
<td>GTOPO30</td>
<td>30</td>
<td>Global*</td>
<td>HYDRO1K</td>
<td>Grid</td>
</tr>
<tr>
<td>GMTED_BASIN</td>
<td>GMTED</td>
<td>SRTM&amp;others</td>
<td>7.5</td>
<td>Global</td>
<td>HYDRO1K</td>
<td>Catchment</td>
</tr>
<tr>
<td>GMTED_GRID</td>
<td>GMTED</td>
<td>SRTM&amp;others</td>
<td>7.5</td>
<td>Global</td>
<td>HYDRO1K</td>
<td>Grid</td>
</tr>
<tr>
<td>SHEDS_BASIN</td>
<td>HydroSHEDS</td>
<td>SRTM</td>
<td>7.5</td>
<td>&lt; 60° N</td>
<td>HydroSHEDS</td>
<td>Catchment</td>
</tr>
<tr>
<td>SHEDS_GRID</td>
<td>HydroSHEDS</td>
<td>SRTM</td>
<td>7.5</td>
<td>&lt; 60° N</td>
<td>HydroSHEDS</td>
<td>Grid</td>
</tr>
</tbody>
</table>
Table 3. Summary of simulated and observed mean annual minimum (MIN), maximum (MAX), and amplitude (AMP) of wetland extent for 1980–2010. All units are Mkm$^2$ ($10^6$ km$^2$) $\pm 1\sigma$, where standard deviation represents the inter-annual variation in model estimates except for the row Average, which represents uncertainties of estimates from each model experiment.

<table>
<thead>
<tr>
<th>Model</th>
<th>Lowland Amazon Basin</th>
<th>West Siberian Lowland</th>
<th>Global</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MIN</td>
<td>MAX</td>
<td>AMP</td>
</tr>
<tr>
<td>SHEDS_BASIN</td>
<td>0.27 ± 0.02</td>
<td>0.38 ± 0.01</td>
<td>0.11 ± 0.01</td>
</tr>
<tr>
<td>SHEDS_GRID</td>
<td>0.32 ± 0.01</td>
<td>0.40 ± 0.01</td>
<td>0.08 ± 0.01</td>
</tr>
<tr>
<td>GMTED_BASIN</td>
<td>0.21 ± 0.02</td>
<td>0.35 ± 0.01</td>
<td>0.14 ± 0.02</td>
</tr>
<tr>
<td>GMTED_GRID</td>
<td>0.19 ± 0.02</td>
<td>0.34 ± 0.01</td>
<td>0.15 ± 0.02</td>
</tr>
<tr>
<td>HYDRO1k_BASIN</td>
<td>0.25 ± 0.02</td>
<td>0.37 ± 0.01</td>
<td>0.12 ± 0.01</td>
</tr>
<tr>
<td>HYDRO1k_GRID</td>
<td>0.22 ± 0.02</td>
<td>0.36 ± 0.01</td>
<td>0.14 ± 0.02</td>
</tr>
<tr>
<td>Average</td>
<td>0.27 ± 0.04</td>
<td>0.38 ± 0.02</td>
<td>0.11 ± 0.01</td>
</tr>
</tbody>
</table>

Observations

- Hess2015: 0.23, 0.58
- GIEMS: 0.12 ± 0.01, 0.25 ± 0.03, 0.14 ± 0.04, 0 ± 0, 0.24 ± 0.05, 0.25 ± 0.05, 1.38 ± 0.09, 4.47 ± 0.20, 3.09 ± 0.19
- SWAMPS-GLWD: 0.22 ± 0.03, 0.34 ± 0.01, 0.12 ± 0.03, 0 ± 0, 0.50 ± 0.03, 0.51 ± 0.03, 3.03 ± 0.13, 6.62 ± 0.18, 3.63 ± 0.14
Table 4. Spearman correlations between satellite-based vs. modeled interannual anomalies of the grid-cells contained in each region defined in Fig. 2f at global scale. Values out and in parentheses are correlation efficient with SWAMPS-GLWD and GIEMS respectively. The two highest value within one column is in bold.

<table>
<thead>
<tr>
<th>Regions</th>
<th>SHDES BASIN</th>
<th>SHDES GRID</th>
<th>GMTED BASIN</th>
<th>GMTED GRID</th>
<th>HYDRO1K BASIN</th>
<th>HYDRO1k GRID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boreal North America</td>
<td>0.770</td>
<td>0.768</td>
<td>0.751</td>
<td>0.745</td>
<td>0.765</td>
<td>0.748</td>
</tr>
<tr>
<td></td>
<td>(0.378)</td>
<td>(0.376)</td>
<td>(0.354)</td>
<td>(0.341)</td>
<td>(0.378)</td>
<td>(0.343)</td>
</tr>
<tr>
<td>Boreal Eurasia</td>
<td>0.785</td>
<td>0.782</td>
<td>0.763</td>
<td>0.764</td>
<td>0.763</td>
<td>0.760</td>
</tr>
<tr>
<td></td>
<td>(0.513)</td>
<td>(0.511)</td>
<td>(0.487)</td>
<td>(0.487)</td>
<td>(0.493)</td>
<td>(0.484)</td>
</tr>
<tr>
<td>Europe</td>
<td>0.604</td>
<td>0.595</td>
<td>0.313</td>
<td>0.211</td>
<td>0.588</td>
<td>0.218</td>
</tr>
<tr>
<td></td>
<td>(0.091)</td>
<td>(0.079)</td>
<td>(−0.198)</td>
<td>(−0.278)</td>
<td>(0.076)</td>
<td>(−0.272)</td>
</tr>
<tr>
<td>Tropical South America</td>
<td>0.723</td>
<td>0.725</td>
<td>0.724</td>
<td>0.666</td>
<td>0.708</td>
<td>0.726</td>
</tr>
<tr>
<td></td>
<td>(0.838)</td>
<td>(0.831)</td>
<td>(0.835)</td>
<td>(0.825)</td>
<td>(0.836)</td>
<td>(0.835)</td>
</tr>
<tr>
<td>South Africa</td>
<td>0.082</td>
<td>0.044</td>
<td>0.084</td>
<td>0.076</td>
<td>0.040</td>
<td>0.088</td>
</tr>
<tr>
<td></td>
<td>(0.736)</td>
<td>(0.725)</td>
<td>(0.735)</td>
<td>(0.734)</td>
<td>(0.717)</td>
<td>(0.740)</td>
</tr>
<tr>
<td>Tropical Asia</td>
<td>0.689</td>
<td>0.681</td>
<td>0.705</td>
<td>0.677</td>
<td>0.670</td>
<td>0.648</td>
</tr>
<tr>
<td></td>
<td>(0.674)</td>
<td>(0.673)</td>
<td>(0.682)</td>
<td>(0.625)</td>
<td>(0.660)</td>
<td>(0.632)</td>
</tr>
<tr>
<td>Temperate North America</td>
<td>0.359</td>
<td>0.380</td>
<td>0.406</td>
<td>0.347</td>
<td>0.518</td>
<td>0.479</td>
</tr>
<tr>
<td></td>
<td>(0.139)</td>
<td>(0.155)</td>
<td>(0.262)</td>
<td>(0.229)</td>
<td>(0.288)</td>
<td>(0.305)</td>
</tr>
<tr>
<td>Temperate South America</td>
<td>−0.193</td>
<td>−0.205</td>
<td>−0.153</td>
<td>−0.162</td>
<td>−0.178</td>
<td>−0.166</td>
</tr>
<tr>
<td></td>
<td>(0.633)</td>
<td>(0.597)</td>
<td>(0.622)</td>
<td>(0.641)</td>
<td>(0.627)</td>
<td>(0.627)</td>
</tr>
<tr>
<td>Temperate Eurasia</td>
<td>0.742</td>
<td>0.760</td>
<td>0.735</td>
<td>0.721</td>
<td>0.732</td>
<td>0.716</td>
</tr>
<tr>
<td></td>
<td>(0.645)</td>
<td>(0.660)</td>
<td>(0.642)</td>
<td>(0.643)</td>
<td>(0.642)</td>
<td>(0.642)</td>
</tr>
</tbody>
</table>
Table A1. Reclassification table for aggregating JERS-1 lowland Amazon basin to 0.5° cell. Code NA, 0, 1, and 2 represent Not-Available, Not Wetlands, wetland only exist in low-water season and wetland exist in high-water season.

<table>
<thead>
<tr>
<th>DN</th>
<th>Cover at low-water stage</th>
<th>Cover at higher-water stage</th>
<th>Flag for minimum/maximum wetlands</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Land outside Amazon Basin</td>
<td>Land outside Amazon Basin</td>
<td>NA</td>
</tr>
<tr>
<td>1</td>
<td>Non-wetland within Amazon Basin</td>
<td>Non-wetland within Amazon Basin</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>Open water</td>
<td>Open water</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>Open water</td>
<td>Aquatic macrophyte</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>Bare soil or herbaceous, non-flooded</td>
<td>Open water</td>
<td>2</td>
</tr>
<tr>
<td>23</td>
<td>Bare soil or herbaceous, non-flooded</td>
<td>Aquatic macrophyte</td>
<td>2</td>
</tr>
<tr>
<td>33</td>
<td>Aquatic macrophyte</td>
<td>Aquatic macrophyte</td>
<td>1</td>
</tr>
<tr>
<td>41</td>
<td>Shrub, non-flooded</td>
<td>Open water</td>
<td>2</td>
</tr>
<tr>
<td>44</td>
<td>Shrub, non-flooded</td>
<td>Shrub, non-flooded</td>
<td>0</td>
</tr>
<tr>
<td>45</td>
<td>Shrub, non-flooded</td>
<td>Shrub, flooded</td>
<td>2</td>
</tr>
<tr>
<td>51</td>
<td>Shrub, flooded</td>
<td>Open water</td>
<td>1</td>
</tr>
<tr>
<td>55</td>
<td>Shrub, flooded</td>
<td>Shrub, flooded</td>
<td>1</td>
</tr>
<tr>
<td>66</td>
<td>Woodland, non-flooded</td>
<td>Woodland, non-flooded</td>
<td>0</td>
</tr>
<tr>
<td>67</td>
<td>Woodland, non-flooded</td>
<td>Woodland, flooded</td>
<td>2</td>
</tr>
<tr>
<td>77</td>
<td>Woodland, flooded</td>
<td>Woodland, flooded</td>
<td>1</td>
</tr>
<tr>
<td>88</td>
<td>Forest, non-flooded</td>
<td>Forest, non-flooded</td>
<td>0</td>
</tr>
<tr>
<td>89</td>
<td>Forest, non-flooded</td>
<td>Forest, flooded</td>
<td>2</td>
</tr>
<tr>
<td>99</td>
<td>Forest, flooded</td>
<td>Forest, flooded</td>
<td>1</td>
</tr>
<tr>
<td>200</td>
<td>Elevation ≤ 500 m, in Basin</td>
<td>Elevation ≤ 500, in Basin</td>
<td>NA</td>
</tr>
<tr>
<td>255</td>
<td>Ocean</td>
<td>Ocean</td>
<td>NA</td>
</tr>
</tbody>
</table>
Figure 1. Cumulative distribution function (CDF) of the fitted exponential curve (blue line) as a function of compound topographic index (CTI) in comparison with the three-parameter gamma function (red line), as well as the observations (grey line) with in a sample grid box.
Figure 2. Fmax, Cs, Mean CTI in LPJ-wsl, and Transcom regions.
Figure 3. Evaluation of permafrost simulation in LPJ-wsl. (a) Inventory-based (light blue) and simulated (dark blue) permafrost extent from NSIDC and LPJ-wsl respectively. The inventory contains discontinuous, sporadic or isolated permafrost boundaries, as well as the location of subsea and relict permafrost. We only compare the distribution of all permafrost against model outputs without distinguishing each permafrost types. (b) Spatial distribution of Spearman correlation between simulated monthly frozen-days from LPJ-wsl over 2002–2011 and satellite retrievals of FT status from AMSRE.
**Figure 4.** Comparison of TOPMODEL-based wetland areas and Observational datasets over the region West Siberian Lowland (WSL) for June–July–August (JJA) average over the period 1993–2012. “Calibrated” and “Original” represent simulated wetland areas with and without $F_{\text{max}}$ calibration respectively. For Sheng2004, Tanocai, Pregon2008, and GLWD, it represents maximum wetland extent per 0.5° cell as derived from static inventory maps. For SWAMPS-GLWD and GIEMS, areas shown are averaged for JJA over the period 1993–2007 and 2000–2012 respectively.
Figure 5. Comparison of wetland areas (km\(^2\)) between LPJ-wsl simulated results (SHEDS_basin version) and JERS-1 satellite observation for low-water season and high-water season. The low water season and high-water season in LPJ was calculated by mean annual minimum and maximum respectively during 1993–2013.
Figure 6. Observation-based estimate from Glagolev et al., 2011 and two LPJ-wsl estimates using Hydro-SHEDS (calibrated $F_{\text{max}}$ and non-calibrated $F_{\text{max}}$) for annual CH4 emission ($\text{g CH}_4 \text{ yr}^{-1} \text{ m}^{-2}$ of grid cell area). Averages from LPJ-wsl are over the time period 2007–2010.
**Figure 7.** Spatial distributions of average June–July–August (JJA) wetland area (km$^2$) over the West Siberian Lowland (WSL) area from model experiments.
**Figure 8.** Average seasonal variation of observed and simulated monthly total wetland area for Transcom regions. For consistent comparison, two sets of simulated results were generated by masking out pixels for which GIEMS (red, dashed) or SWAMPS-GLWD (blue, dashed) do not have observations (denoted as “-G” and “-S”, respectively).
Figure 9. Global wetland potential map, which is calculated by the ratio of the mean annual maximum wetland extent averaged for the time period 1980–2010 and the long-term potential maximum wetland area ($F_{\text{wet max}}^\text{wet}$). Higher value represents higher availability for sub-grids to be inundated.
Figure A1.
Figure A1. Interannual variations of seasonal wetland area anomalies from LPJ-wsl and satellite-derived observations for the period 1993–2012.