NACP MsTMIP: Global and North American Driver Data for Multi-Model Intercomparison

Get Data

Revision date: July 8, 2014

Summary:

This data set provides environmental data that have been standardized and aggregated for use as input to carbon cycle models at global (0.5-degree resolution) and regional (North America at 0.25-degree resolution) scales.

The data were compiled from selected sources (Table 2) and integrated into gridded global and regional collections of climatology variables (precipitation, air temperature, air specific humidity, air relative humidity (NA only), pressure, downward longwave radiation, downward shortwave radiation, and wind speed), time-varying atmospheric CO2 concentrations, time-varying nitrogen deposition, biome fraction and type, land-use and land-cover change, C3/C4 grasses fractions, major crop distribution, phenology, multiple soil characteristics, and a land-water mask. The driver data are sufficient for carbon cycle model simulations from 1801 to 2010. The temporal resolution ranges from 3-hourly for climate to monthly for CO2-atm and phenology to annual for N-deposition and landcover.

These data were compiled specifically for the North American Carbon Program (NACP) Multi-Scale Synthesis and Terrestrial Model Intercomparison Project (MsTMIP) as the prescribed model input driver data (Huntzinger et al., 2013). The driver data were used by 22 terrestrial biosphere models to run baseline and sensitivity simulations. The standardized data provided consistent model inputs to minimize the inter-model variability caused by differences in environmental drivers and initial conditions. Together with the sensitivity simulations, the standardized input data enable better interpretation and quantification of structural and parameter uncertainties of model estimates.

Data are provided in Climate and Forecast (CF) metadata convention compliant (version 1.4) netCDF-4 file formats. There are 3,152 *.nc4 data files with this data set.

The compilation of these data was facilitated by the NACP Modeling and Synthesis Thematic Data Center (MAST-DC). MAST-DC was a component of the NACP (www.nacarbon.org) designed to support NACP by providing data products and data management services needed for modeling and synthesis activities. The overall objective of MAST-DC was to provide data management support to NACP investigators and agencies performing modeling and synthesis activities.

Data and Documentation Access:

Get Data: http://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1220

Companion Documentation for this Data Set:

A copy of this user's guide is included as a companion file.

Related Data Products:

MsTMIP Model Structure data

MsTMIP Model Output data (to be archived)

Other Model Driver Data Archived at the ORNL DAAC:

VEMAP 2: U.S. Monthly Climate, 1895-1993, Version 2

The International Satellite Land Surface Climatology Project, Initiative II (ISLSCP II)

Data Citation:

Cite this data set as follows:

Wei, Y., S. Liu, D.N. Huntzinger, A.M. Michalak, N. Viovy, W.M. Post, C.R. Schwalm, K. Schaefer, A.R. Jacobson, C. Lu, H. Tian, D.M. Ricciuto, R.B. Cook, J. Mao, and X. Shi. 2014. NACP MsTMIP: Global and North American Driver Data for Multi-Model Intercomparison. Data set. Available on-line [http://daac.ornl.gov] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA. http://dx.doi.org/10.3334/ORNLDAAC/1220

Table of Contents:

1. Data Set Overview

Project: North American Carbon Program (NACP)

The NACP (Denning et al., 2005; Wofsy and Harriss, 2002) is a multidisciplinary research program to obtain scientific understanding of North America's carbon sources and sinks and of changes in carbon stocks needed to meet societal concerns and to provide tools for decision makers. Successful execution of the NACP has required an unprecedented level of coordination among observational, experimental, and modeling efforts regarding terrestrial, oceanic, atmospheric, and human components. The project has relied upon a rich and diverse array of existing observational networks, monitoring sites, and experimental field studies in North America and its adjacent oceans. It is supported by a number of different federal agencies through a variety of intramural and extramural funding mechanisms and award instruments. The Oak Ridge National Laboratory (ORNL) Distributed Active Archive Center (DAAC) is the archive for the NACP data products.

Data Products (Wei et al. (in review))

The variables in this data set include climatology [precipitation, air temperature, air specific humidity, air relative humidity (NA only), pressure, downward longwave radiation, downward shortwave radiation, and wind speed], time-varying atmospheric CO2 concentrations, time-varying nitrogen deposition, biome fraction and type, land-use and land-cover change (LULCC), C3/C4 grasses fractions, major crop distribution, phenology, soil characteristics data, and a land-water mask (Table 2). The data are provided in a standard format for global (0.5 degree by 0.5 degree resolution) and regional (North American, 0.25 degree by 0.25 degree resolution) analyses.

For most data categories, the North American data sets are based on the same data sources as the global products. However, different climatology and soil data products were compiled for the two domains based primarily on the availability of these data at the spatial and temporal resolution needed for the model simulations. In order to meet the needs of MsTMIP, improvements were made to the quality and/or the spatial and temporal coverage and resolution of several of the original environmental data sets (Wei et al. (in review)).

The resulting standardized data are provided in Climate & Forecast (CF) 1.4 convention compliant netCDF version 4 format, which is supported by a wide range of programming APIs (e.g., C, C++, Fortran, Java, Perl) and multiple operating systems (e.g., Linux, Unix, Mac OS X, Windows). All drivers are saved in Coordinated Universal Time (UTC) with all sub-monthly drivers (e.g., climate) including leap years.

MsTMIP -- Model Intercomparison

This data set provided standardized environmental driver data for 22 terrestrial biosphere models (TBM) participating in the NACP Multi-scale Synthesis and Terrestrial Model Intercomparison Project (MsTMIP). MsTMIP is a formal multi-scale and multi-model intercomparison and evaluation effort focused on improving the diagnosis and attribution of carbon exchange at regional and global scales (Huntzinger et al., 2013; Wei et al., in review). MsTMIP builds upon current and past NACP synthesis activities, and was established to provide a consistent and unified modeling framework to isolate, interpret, and inform understanding of how TBM structural differences and associated internal parameters impact estimates of terrestrial ecosystem carbon dynamics. [Model structure refers to the types of processes considered (e.g., nutrient cycling, disturbance, lateral transport of carbon), and how these processes are represented (e.g., photosynthetic formulation, temperature sensitivity, respiration parameterization) in the models.] By prescribing a common experimental protocol (Huntzinger et al., 2013) with standardized driver data (this data set) and spin-up procedures for all model simulations (Wei et al., in review), the biases and variability in TBM estimates of regional and global carbon budgets resulting from differences in the models themselves and model-specific parameter values can be isolated and explained.

Data Synthesis Process (Wei et al. (in review))

Working closely with the core MsTMIP team, MAST-DC prepared the driver data to the requirements needed by the MsMTIP protocol and the individual modeling teams (20+). Based on project requirements, MAST-DC compiled climate, atmospheric CO2 concentrations, nitrogen deposition, land-use and land-cover change (LULCC), C3/C4 grasses fractions, major crops, phenology, and soil data from identified sources into a standard format for global (0.5 degree by 0.5 degree resolution) and regional (North America, 0.25 degree by 0.25 degree resolution) model simulations. In order to meet the needs of MsTMIP, improvements were made to several of the original environmental data sets, by changing the quality, the spatial and temporal coverage, resolution, or a combination of these as described in Section 5.

Table 1. Author's afilliation and role in development of MsTMIP driver data

Name	Institution	Email	Role
Yaxing Wei	Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA	weiy@ornl.gov	MAST-DC
Shishi Liu	Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA	carol.shishi@gmail.com	MAST-DC
Deborah H. Huntzinger	School of Earth Sciences and Environmental Sustainability & the Department of Civil Engineering, Construction Mgmt., and Environmental Engineering, Northern Arizona University, Flagstaff, AZ, USA	deborah.huntzinger@nau.edu	Core MsTMIP
Anna Michalak	Department of Global Ecology, Carnegie Institution for Science, Stanford, CA, USA	michalak@stanford.edu	Core MsTMIP
Viovy Nicolas	Laboratoire des Sciences du Climat et l'Environnement, Paris, France	viovy@lsce.ipsl.fr	MsTMIP
Wilfred M. Post (retired)	Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA	mpost3116@gmail.com	MAST-DC
Christopher Schwalm	School of Earth Sciences and Environmental Sustainability & the Department of Civil Engineering, Construction Mgmt., and Environmental Engineering, Northern Arizona University, Flagstaff, AZ, USA	Christopher.Schwalm@nau.edu	Core MsTMIP
Kevin Schaefer	University of Colorado, National Snow and Ice Data Center, Boulder, CO, USA	kevin.schaefer@nsidc.org	Core MsTMIP
Andrew R. Jacobson	University of Colorado and NOAA Earth System Research Lab, Boulder, CO, USA	andy.jacobson@noaa.gov	Core MsTMIP
Chaoqun Lu	International Center for Climate and Global Change Research, School of Forestry and Wildlife Sciences, Auburn University, Auburn, AL, USA	czl0003@auburn.edu	Modeler
Hanqin Tian	International Center for Climate and Global Change Research, School of Forestry and Wildlife Sciences, Auburn University, Auburn, AL, USA	tianhan@auburn.edu	Modeler
Dan M. Ricciuto	Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA	ricciutodm@ornl.gov	Modeler
Robert B. Cook	Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA	cookrb@ornl.gov	MAST-DC
Jiafu Mao	Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA	maoj@ornl.gov	Modeler
Xiaoying Shi	Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA	shix@ornl.gov	Modeler

Table 2. The MsTMIP Driver Data Summary

Category	Data Source	Spatial Extent & Resolution (degrees)	Native Temporal Period, Resolution	Extended Temporal Period, Resolution*	Variables (Units)
Climate	CRUNCEP	Global (0.5)	1901-2010, 6-hourly	1801-2010, 6-hourly	Total Precipitation (mm) Air temperature Air specific humidity (g/g) Air relative humidity (NA only) Pressure (Pa) Downward longwave radiation (W/m2) Downward shortwave radiation Wind speed

	NARR	NA (0.25)	1979-2010, 3-hourly	1801-2010, 3-hourly	
Land Water Mask	CRUNCEP	Global (0.5)	Constant	Constant	Binary land vs. water map; 1 represents land; 0 represents water
	NARR	NA (0.25)			
CO2	Extended GLOBALVIEW-CO2	Global (0.5) , NA (0.25)	1700-2010, monthly	1801-2010, monthly	Atmospheric CO2 concentration
Nitrogen Deposition	Enhanced Dentener	Global (0.5) , NA (0.25)	1860-2050, annual	1801-2010, annual	NHx-N deposition NOy-N deposition
Biome	SYNMAP	Global (0.5) , NA (0.25)	Constant Present = around year 2000 Potential = pre-industry, pre-agriculture	Constant	Biome fraction Biome type
Land-Use and Land-Cover Change	SYNMAP + Hurtt	Global (0.5) , NA (0.25)	1700-2010, annual	1801-2010, annual	Biome fraction Biome type Biome type pure
C3 /C4 Grass	C3/C4 grass fraction	Global (0.5) , NA (0.25)	Constant Present = around year 2000 Potential = pre-industry, pre-agriculture (around year 1900 or before)	Constant	Present relative fractions of C3/C4 grasses Potential relative fractions of C3/C4 grasses
Major Crops	Monfreda et al. (2008)	Global (0.5) , NA (0.25)	Constant (year 2000)	Constant	Fraction of harvest area in each grid cell for maize, rice, soybean, and wheat
Phenology	GIMMSg	Global (0.5) , NA (0.25)	1700-2010, monthly	1801-2010, monthly	NDVI, LAI, fPAR
Soil	HWSD v1.1	Global (0.5)	Constant	Constant	Soil layers Dominant soil type Reference soil depth Clay/sand/silt fractions pH Organic carbon Cation exchange capacity Reference bulk density Gravel content
	Unified North American Soil Database (UNASD) [STATSGO2 (US) + SLC v3.2 and v2.2 (CA) + HWSD 1.1]	NA (0.25)			

Notes:

* Native temporal periods of environmental driver data sets were extended for the TBM simulation time period (1801-2010) defined by MsTMIP. Please refer to Data Acquisition Materials and Methods section to see how a data with shorter native temporal period are extended back to 1801 to address the needs of MsTMIP simulations.

CRUNCEP = Climate Research Unit, National Centers for Environmental Prediction

NARR = North American Regional Reanalysis

SYNMAP = SYNergetic land cover MAP

GIMMSg = Global Inventory Monitoring and Modeling System version g

NDVI = Normalized Difference Vegetation Index

LAI = Leaf Area Index

fPAR = canopy absorbed fraction of Photosynthetically Active Radiation

HWSD = Harmonized World Soil Database

STATSGO2 = State Soil Geographic data version 2

SLC = Soil Landscapes of Canada

2. Data Characteristics:

2.1. Spatial Coverage

Sites: Global and North America

Site Boundaries:(All latitude and longitude given in decimal degrees)

Site (Region)	Westernmost Longitude	Easternmost Longitude	Northernmost Latitude	Southernmost Latitude
Global (all land surface areas excluding Antarctica)	-178.750	179.950	89.750	-78.250
North America	-169.875	-50.125	83.875	10.125

2.2. Spatial Resolution

Environmental model driver data are provided globally at 0.5 degree by 0.5 degree resolution and regionally over North America at 0.25 degree by 0.25 degree resolution.

2.3. Temporal Coverage

Environmental model driver data are provided in their respective "native" temporal period for both global and North America resolutions as shown in Table 2. The temporal ranges of the data were modified to enable carbon cycle model simulations from 1801 to 2010.

2.4. Temporal Resolution

For model simulations, global climate data are 6-hourly and North American climate data are 3-hourly. For both global and North American model simulations, CO2 concentrations and phenology are monthly, nitrogen deposition and land use and cover change are annual, and relative fractions of C3/C4 grasses, major crops, soils, and land/sea mask are constant.

All driver data are saved in Coordinated Universal Time (UTC) with all sub-monthly drivers (e.g., climate) including leap years. For models using an internal time step less than the driver data time step, the model linearly interpolated between weather data points, except for the down-welling shortwave radiation, where scaling using the cosine of the zenith angle was appropriate. If a model used a time step larger than the drivers, appropriate time averages or totals of the weather data were use. For example, a model with a 1-day time step would use 24-hour averages or totals.

The sub-daily climate data include leap years. If a model did not account for leap year, the MAST-DC removed February 29 from the driver data in leap years. They did not delete December 31, January 1, or any other day because this would have created a time lag between model output and the observations.

2.5. Data File Information

All MsTMIP model driver data files, including global data and North American data, are stored in CF metadata convention compliant (version 1.4) netCDF-4 file format. Climate variables, which have a fine temporal resolution, are provided in multiple netCDF files with each file containing one year of data. Other variables with a coarser temporal resolution, including CO2 concentration, phenology, nitrogen deposition, and land use change, are supplied in individual netCDF files covering the whole temporal period.

Missing Value: -999.0.

File naming convention:

All data files are named **mstmip_driver_XX_ZZ_variable____Year_v1.nc4**

where

XX=global or na (North America), **ZZ**=hd (half degree) or qd (quarter degree)

User Note: Be advised that all 3,152 files are listed in a flat structure. Also, in the tables below, "mstmip_driver" has been omitted from the file names in order to save space in the tables.

MsTMIP DATA FILES

-

Land-water Masks (two data files)

Global CRUNCEP Land Mask:
Spatial Resolution/Extent: 0.5 degree / W: -180, S: -90, E: 180, N: 90
Temporal Resolution/Extent: One time

North America NARR Land Mask:
Spatial Resolution/Extent: 0.25 degree / W: -170, S: 10, E: -50, N: 84
Temporal Resolution/Extent: One time

Table 3. Land-water Mask data files

Data Files	Description	Units
mstmip_driver_global_hd_landwatermask_v1.nc4	Global land-water mask at 0.5 degree	1 = land 0 = water
mstmip_driver_na_qd_landwatermask_v1.nc4	North American land-water mask at 0.25 degrees	

Climate Data (1,874 data files)

There are 1,650 global climate data files and 224 data files for North America (na). The files are described in tables 4 and 5 below.

Global Climate (CRU+NCEP)

Spatial Resolution/Extent: 0.5 degree / W: -180, S: -90, E: 180, N: 90
Temporal Resolution/Extent: 6-hourly / 1901-2010
Number of Files for Each Variable: 110 annual files

Table 4. Global Climate (CRU+NCEP) data files

Variable	Example file name	Description	Units
lwdown	_global_hd_climate_lwdown_1901_v1.nc4	6 hourly incoming longwave radiation	W/m2
press	_global_hd_climate_press_1901_v1.nc4	6 hourly pressure	Pa
press_monthly_mean	_global_hd_climate_press_monthly_mean_2010_v1.nc4	Monthly mean pressure	
qair	_global_hd_climate_qair_2008_v1.nc4	6 hourly air specific humidity	g/g
qair_monthly	_global_hd_climate_qair_monthly_mean_2010_v1.nc4	Monthly mean air specific humidity	
rain	_global_hd_climate_rain_1901_v1.nc4	6 hourly total rainfall	mm
rain_total_monthly	_global_hd_climate_rain_monthly_total_2010_v1.nc4	Monthly total rainfall	
swdown	_global_hd_climate_swdown_1978_v1.nc4	6 hourly average incoming shortwave radiation	W/m2
swdown_monthly_mean	_global_hd_climate_swdown_monthly_mean_2006_v1.nc4	Monthly mean incoming shortwave radiation	W/m2
swdown_total_6hourly	_global_hd_climate_swdown_6hourly_total_1961_v1.nc4	6 hourly total incoming shortwave radiation energy	J/m2
swdown_monthly_total	_global_hd_climate_swdown_monthly_total_1072_v1.nc4	Monthly total incoming shortwave radiation energy	J/m2
tair	_global_hd_climate_tair_1914_v1.nc4	6 hourly instantaneous air temperature at 2m	K
tair_monthly_mean	_global_hd_climate_tair_monthly_mean_1938_v1.nc4	6 hourly monthly mean air temperature	
uwind	_global_hd_climate_uwind_1978_v1.nc4	6 hourly wind speed of u wind component	m/s
vwind	_global_hd_climate_vwind_1978_v1.nc4	6 hourly wind speed of v wind component	

North American Climate (NARR)

Spatial Resolution/Extent: 0.25 degree / W: -170, S: 10, E: -50, N: 84
Temporal Resolution/Extent: 3-hourly / 1979-2010
Number of Files for Each Variable: 32 annual files

Table 5. North American Climate (NARR)

Variable	Example File Name	Description	Cell Methods	Units
air_2m	_na_qd_climate_air_2m_1987_v1.nc4	Air temperature at 2m[1]	Spatial: average Temporal: instantaneous value at start point of each 3-hour period	K
apcp	_na_qd_climate_apcp_1987_v1.nc4	Accumulated total precipitation at surface[1]	Spatial: average Temporal: sum in a 3-hour period	kg/m2
dlwrf	_na_qd_climate_dlwrf_2000_v1.nc4	Downward longwave radiation flux at surface[1]	Spatial: average Temporal: average in a 3-hour period	W/m2
dswrf	_na_qd_climate_dswrf_2000_v1.nc4	Downward shortwave radiation flux at surface[2]	Spatial: average Temporal: average in a 3-hour period	W/m2
rhum_2m	_na_qd_climate_rhum_2m_2000_v1.nc4	Relative humidity at 2m	Spatial: average Temporal: instantaneous value at start point of each 3-hour period	%
shum_2m	_na_qd_climate_shum_2m_2000_v1.nc4	Specific humidity at 2m[1]	Spatial: average Temporal: instantaneous value at start point of each 3-hour period	kg/kg
wnd_10m	_na_qd_climate_wnd_10m_1996_v1.nc4	Wind Speed at 10m[3]	Spatial: average Temporal: instantaneous value at start point of each 3-hour period	m/s

Notes:

[1]Regridded to quarter degree resolution using distance-weighted average method.

[2]Regridded to quarter degree resolution using area-weighted average method, rescaled with analysis output from MTCLIM algorithm. Rows with missing values replaced with nearby values. Negative values replaced with 0.

[3]Wind speed was calculated from U and V wind components, then regridded to quarter degree resolution using distance-weighted average method.

Atmospheric CO2 Concentration (two data files)

Global:
Spatial Resolution/Extent: 0.5 degree / W: -180, S: -90, E: 180, N: 90
Temporal Resolution/Extent: Monthly / 1700 - 2010

North America:
Spatial Resolution/Extent: 0.25 degree / W: -170, S: 10, E: -50, N: 84
Temporal Resolution/Extent: Monthly / 1700 - 2010

Table 6. Atmospheric CO2 Concentration data files

FILES	DESCRIPTION	UNITS
_global_hd_co2_v1.nc4	Global CO2 concentration	ppm
_na_qd_co2_v1.nc4	North American CO2 concentration	

Nitrogen Deposition (four data files)

Global:
Spatial Resolution/Extent: 0.5 degree / W: -180, S: -90, E: 180, N: 90
Temporal Resolution/Extent: Annual / 1860 - 2050

North America:
Spatial Resolution/Extent: 0.25 degree / W: -170, S: 10, E: -50, N: 84
Temporal Resolution/Extent: Annual / 1860 - 2050

Table 7. Nitrogen Deposition data files

FILES	DESCRIPTION	UNITS
_global_hd_nitrogen_nhx_v1.nc4	Global NHx-N deposition at half degrees	mgN/m2/yr
_global_hd_nitrogen_noy_v1.nc4	Global NOy-N deposition at half degrees	
_na_qd_nitrogen_nhx_v1.nc4	North American NHx-N deposition at 0.25 degrees	
_na_qd_nitrogen_noy_v1.nc4	North American NOy-N deposition at 0.25 degrees	

Biome Data (two data files)

There are two global biome data files and two files for North America. The files are described in tables 8 and 9 below.

Global SYNMAP BIOME data:

Present SYNMAP:
Spatial Resolution/Extent: 0.5 degree / W: -180, S: -90, E: 180, N: 90
Temporal Resolution/Extent: one time (around 2000)

Potential SYNMAP:
Spatial Resolution/Extent: 0.5 degree / W: -180, S: -90, E: 180, N: 90
Temporal Resolution/Extent: one time (pre-industry, pre-agriculture)

Table 8. SYNMAP Biome Classification data files

FILES	DESCRIPTION	VARIABLE NAME (definition)	UNITS
_global_hd_biome_v1.nc4	Present global SYNMAP	biome_frac (fraction of each biome type in each cell)	fraction
		biome_type (dominant biome type in each cell)	none
_global_hd_biome_potveg_v1.nc4	Potential global SYNMAP	biome_frac (fraction of each biome type in each cell)	fraction
		biome_type (dominant biome type in each cell)	none

North American SYNMAP Biome Classification (two data files)

Present SYNMAP:
Spatial Resolution/Extent: 0.25 degree / W: -170, S: 10, E: -50, N: 84
Temporal Resolution/Extent: one time (around 2000)

Potential SYNMAP:
Spatial Resolution/Extent: 0.25 degree / W: -170, S: 10, E: -50, N: 84
Temporal Resolution/Extent: one time (pre-industry, pre-agriculture)

Table 9. North American SYNMAP Biome Classification data files

Data Files	Description	Variable name(definition)	Units
_na_hd_biome_v1.nc4	Present North American SYNMAP	biome_frac (fraction of each biome type in each cell)	fraction
		biome_type (dominant biome type in each cell)	none
_na_hd_biome_potveg_v1.nc4	Potential North American SYNMAP	biome_frac (fraction of each biome type in each cell)	fraction
		biome_type (dominant biome type in each cell)	none

Land-Use and Land-Cover Change (Hurtt_SYNMAP data)

There are 622 data files of land use and land cover change; 311 global data files and 311 files for North America. The variables and files are described in the table below.

Global:

Spatial Resolution/Extent: 0.5 degree / W: -180, S: -90, E: 180, N: 90
Temporal Resolution/Extent: Annual / 1700 - 2010

North America:
Spatial Resolution/Extent: 0.25 degree / W: -170, S: 10, E: -50, N: 84
Temporal Resolution/Extent: Annual / 1700 - 2010

Table 10. Land-Use and Land-Cover Change data files

Variable	Variable definition	Example file names	Description	Units
land_use_change (lulcc)	biome_frac	_global_hd_lulcc_1776_v1.nc4 _na_qd_lulcc_1776_v1.nc4	A 3 D array with dimensions latitude, longitude and merged time - varying fractional coverages for all 48 SYNMAP land cover classes. There are 311 files, one for each year.	fraction
	biome_type_pure[1]		A 3D array with dimensions latitude, longitude and time with dominant class only.	none
	biome_type[2]		A 3D array with dimensions latitude, longitude and time with dominant class only.	none

Note: Hurtt_SYNMAP are yearly land-use and land-cover change data created by MsTMIP by merging a static satellite-based land cover product, SYNMAP, with the time-varying land use harmonization (LUH) data for the fifth Assessment Report (AR5) of the Intergovernmental Panel on Climate Change (IPCC). See Data Acquisition Materials and Methods section for details.

[1]Dominant class was determined by first apportioning each SYNMAP mixed class (e.g., Shrubs & Grasses) to a pure class. That is, mixed classes were assumed to exhibit a 50 - 50 split and these fractional coverages were moved to their pure class analogues. After this step the dominant class was given by the pure SYNMAP class with the highest fractional coverage.

[2]Dominant class was given by the SYNMAP class with the highest fractional coverage.

The SYNMAP land cover types are denoted as SYNMAP-Legend:

Value	Life forms	Tree leaf type	Tree leaf longevity	% land
1	Trees	Needle	Evergreen	9.8
2	Trees	Needle	Deciduous	1.7
3	Trees	Needle	Mixed	0.6
4	Trees	Broad	Evergreen	8.2
5	Trees	Broad	Deciduous	3
6	Trees	Broad	Mixed	0.5
7	Trees	Mixed	Evergreen	0.25
8	Trees	Mixed	Deciduous	0.03
9	Trees	Mixed	Mixed	3.5
10	Trees & Shrubs	Needle	Evergreen	1.5
11	Trees & Shrubs	Needle	Deciduous	0.2
12	Trees & Shrubs	Needle	Mixed	0.05
13	Trees & Shrubs	Broad	Evergreen	0.3
14	Trees & Shrubs	Broad	Deciduous	2.5
15	Trees & Shrubs	Broad	Mixed	0.2
16	Trees & Shrubs	Mixed	Evergreen	0.03
17	Trees & Shrubs	Mixed	Deciduous	0.04
18	Trees & Shrubs	Mixed	Mixed	0.3
19	Trees & Grasses	Needle	Evergreen	0.2
20	Trees & Grasses	Needle	Deciduous	0.03

21	Trees & Grasses	Needle	Mixed	0.01
22	Trees & Grasses	Broad	Evergreen	0.3
23	Trees & Grasses	Broad	Deciduous	2.15
24	Trees & Grasses	Broad	Mixed	0.15
25	Trees & Grasses	Mixed	Evergreen	0.005
26	Trees & Grasses	Mixed	Deciduous	0.02
27	Trees & Grasses	Mixed	Mixed	0.2
28	Trees & Crops	Needle	Evergreen	0.3
29	Trees & Crops	Needle	Deciduous	0.006
30	Trees & Crops	Needle	Mixed	0.003
31	Trees & Crops	Broad	Evergreen	0.7
32	Trees & Crops	Broad	Deciduous	1.1
33	Trees & Crops	Broad	Mixed	0.2
34	Trees & Crops	Mixed	Evergreen	0.01
35	Trees & Crops	Mixed	Deciduous	0.01
36	Trees & Crops	Mixed	Mixed	0.4
37	Shrubs	-	-	4.5
38	Shrubs & Grasses	-	-	8.3
39	Shrubs & Crops	-	-	0.4
40	Shrubs & Barren	-	-	10.5
41	Grasses	-	-	8.3
42	Grasses & Crops	-	-	1.5
43	Grasses & Barren	-	-	0.3
44	Crops	-	-	10.7
45	Barren	-	-	11.7
46	Urban	-	-	0.2
47	Snow & Ice	-	-	5.2
0 = water				

C3 and C4 Grass Fraction (12 data files)

There are six global and six NA data files described in the table below.

Global:
Spatial Resolution/Extent: 0.5 degree / W: -180, S: -90, E: 180, N: 90
Temporal Resolution/Extent: One time
Present: around year 2000
Potential: around year 1900 or before

North America:
Spatial Resolution/Extent: 0.25 degree / W: -170, S: 10, E: -50, N: 84
Temporal Resolution/Extent: One time
Present: around year 2000
Potential: around year 1900 or before

Table 11. C3 and C4 Grass Fraction data files

Data Files	Description	Units
_global_hd_C3_rfrac_potveg_v1.nc4	Potential relative fraction of C3 grasses[1]	fraction
_global_hd_C4_rfrac_potveg_v1.nc4	Potential relative fraction of C4 grasses[1]	
_global_hd_grass_frac_potveg_v1.nc4	Potential relative fraction of grasses[1]	
_global_hd_C3_rfrac_presentveg_v1.nc4	Present relative fraction of C3 grasses[2]	

_global_hd_C4_rfrac_presentveg_v1.nc4	Present relative fraction of C4 grasses[2]	
_global_hd_grass_frac_presentveg_v1.nc4	Present relative fraction of grasses[2]	
_na_qd_C3_rfrac_potveg_v1.nc4	Potential relative fraction of C3 grasses[1]	fraction
_na_qd_C4_rfrac_potveg_v1.nc4	Potential relative fraction of C4 grasses[1]	
_na_qd_grass_frac_potveg_v1.nc4	Potential relative fraction of grasses[1]	
_na_qd_C3_rfrac_presentveg_v1.nc4	Present relative fraction of C3 grasses[2]	
_na_qd_C4_rfrac_presentveg_v1.nc4	Present relative fraction of C4 grasses[2]	
_na_qd_C4_rfrac_presentveg_v1.nc4	Present relative fraction of grasses[2]	

Notes:

[1]based on the potential SYNMAP.

[2]based on the present SYNMAP.

-

Major Crop Types (eight data files)

There are four crop global data files and four crop data files for North America described below. The crop data includes maize, rice, soybean, and wheat.

Global:
Spatial Resolution/Extent: 0.5 degree / W: -180, S: -90, E: 180, N: 90
Temporal Resolution/Extent: One time / 2000

North America:
Spatial Resolution/Extent: 0.25 degree / W: -170, S: 10, E: -50, N: 84
Temporal Resolution/Extent: One time / 2000

Table 12. Major Crop Types data files

Files	Description	Units
_global_hd_crop_frac_maize_v1.nc4	Global maize fraction	fraction of harvest area in each cell[1]
_global_hd_crop_frac_rice_v1.nc4	Global rice fraction	
_global_hd_crop_frac_soybean_v1.nc4	Global soybean fraction	
_global_hd_crop_frac_wheat_v1.nc4	Global wheat fraction	
_na_qd_crop_frac_maize_v1.nc4	North American maize fraction	
_na_qd_crop_frac_rice_v1.nc4	North American rice fraction	
_na_qd_crop_frac_soybean_v1.nc4	North American soybean fraction	
_na_qd_crop_frac_wheat_v1.nc4	North American wheat fraction	

Notes: [1]Some of the cells have fraction values greater than 1, because the crops in these cells were harvested more than once each year. Therefore, the harvest area can be greater than the physical cell area.

-

Phenology (622 data files)

There are 311 global phenology data files and 311 phenology data files for North America. The data include leaf area index (LAI), normalized difference vegetation index (NDVI), and the canopy absorbed fraction of photosynthetically active radiation (fPAR).

Global:
Spatial Resolution/Extent: 0.5 degree / W: -180, S: -90, E: 180, N: 90
Temporal Resolution/Extent: monthly / 1700 - 2010

North America:
Spatial Resolution/Extent: 0.25 degree / W: -170, S: 10, E: -50, N: 84
Temporal Resolution/Extent: monthly / 1700 - 2010

Table 13. Phenology data files

Variable	Example file names	Units
LAI (Leaf area index)	_global_hd_phenology_1700_v1.nc4	m2/m2
NDVI (Normalized Difference Vegetation Index)		unitless
fPAR (canopy absorbed fraction of Photosynthetically Active Radiation)	_na_qd_phenology_1700_v1.nc4	unitless

Note: For both global and NA, GIMMSg phenology data were harmonized with SYNMAP to provide LAI and NDVI values for each of the SYNMAP PFTs appearing in a 0.5 or 0.25 degree cell.

Soil Properties (two data files)

There are two soil data files and 22 soil attributes described in the tables below.

Global:
Spatial Resolution/Extent: 0.5 degree / W: -180, S: -90, E: 180, N: 90
Temporal Resolution/Extent: One time
Source: HWSD version 1.1

North America:
Spatial Resolution/Extent: 0.25 degree / W: -170, S: 10, E: -50, N: 84
Temporal Resolution/Extent: One time
Source: Unified North American Soil Database (UNASD) [U.S. General Soil Map (STATSGO2) + Soil Landscapes of Canada v3.2 and v2.2 + HWSD v1.1]

Table 14. Soil Properties data files

Files	Description
mstmip_driver_global_hd_soil_v1.nc4	Gap-filled global soil data
mstmip_driver_na_hd_soil_v1.nc4	Gap-filled North American soil data

Table 15. Soil Attributes

Soil Property	Description	Units in HWSD v 1.1	Units in UNASD
soil_code	soil mapping unit code	code	code
ref_depth	reference soil depth	code	code
roots	obstacles to roots (Europe only)	code	NA
il	impermeable layer (Europe only)	code	NA
t_cec_clay	topsoil CEC (clay)	cmol/kg	meq/100g
t_clay	topsoil clay fraction	% weight	% weight
t_gravel	topsoil gravel content	% volume	% volume
t_oc	topsoil organic carbon	% weight	% weight
t_ph_h20	topsoil pH (H2O)	-log(H+)	-log(H+)
t_ref_bulk	topsoil bulk density	kg/dm^3	g/cm^3
t_sand	topsoil sand fraction	% weight	% weight
t_silt	topsoil silt fraction	% weight	% weight
t_usda_tex	topsoil USDA texture classification	name	name
s_cec_clay	subsoil CEC (clay)	cmol/kg	meq/100g

s_clay	subsoil clay fraction	% weight	% weight
s_gravel	subsoil gravel content	% volume	% volume
s_oc	subsoil organic carbon	% weight	% weight
s_ph_h20	subsoil pH (H2O)	-log(H+)	-log(H+)
s_ref_bulk	subsoil bulk density	kg/dm^3	g/cm^3
s_sand	subsoil sand fraction	% weight	% weight
s_silt	subsoil silt fraction	% weight	% weight
s_usda_tex	subsoil USDA texture classification	name	name

2.6. Companion File Information

A copy of this user's guide is included as a companion file.

3. Data Application and Derivation:

This data product contributes to a multidisciplinary research program to obtain scientific understanding of North America's carbon sources, carbon sinks, and changes in carbon stocks. This information is needed to meet societal concerns and to provide tools for decision makers.

Multi-model intercomparison projects (MIPs) help to characterize or synthesize current understanding of land-atmosphere carbon exchange, and inform the uncertainty or confidence surrounding projections of future exchange and feedbacks with the climate system. The previous NACP MIP conducted on a regional scale provided a comprehensive assessment of the range of estimates of land-atmosphere carbon exchange and uncertainties associated with such estimates, including uncertainties resulting not only from model formulation and assumptions, but also from the choice of environmental driver data and spin up procedures (Huntzinger et. al. 2012) . However, the lack of consistent forcing data and detailed simulation protocols have precluded the attribution of observed across-model variability to differences in modeling approaches. The goal of MsTMIP is to quantify, within a unified intercomparison framework, the contribution of model structural differences to across-model variability in estimates of land-atmosphere carbon exchange, thus providing the critical synthesis, benchmarking, evaluation, and feedback needed to improve the current state of the art in carbon cycle modeling. The MsTMIP experimental protocol (Huntzinger et al., 2013) specifies standard model inputs (this data set), simulations and simulation setup procedures, as well as required model output and format to ensure a valid and fair comparison of model results against one another and against available observations. Wei et al. (in review) outline key components of the MsTMIP environmental model driver data set.

4. Quality Assessment:

Quality analyses of MsTMIP model driver data were conducted by MAST-DC in close collaboration with the MsMTIP team as described by Wei et al. (in review). These quality analyses are summarized below:

- MAST-DC combined the strengths of the CRU and NCEP/NCAR Reanalysis gridded global climatology data sets and corrected known biases in temperature and shortwave radiation in the NCEP/NCAR Reanalysis product. A comparison between the zonal mean of long-term mean downward shortwave radiation for each 0.5 degrees grid cell over land shows that CRU-NCEP has lower downward shortwave radiation than the original NCEP/NCAR data, except at 0 degrees to 10 degrees North and 50 degrees to 55 degrees South, where CRU-NCEP downward shortwave radiation is similar to NCEP/NCAR Reanalysis.
- In order to remove biases in the NARR North American precipitation values (Sun and Barros, 2010), MAST-DC rescaled the NARR 3-hourly precipitation using the GPCP v2.1 (Adler et al., 2003) which has a correction to compensate for systematic biases in gauge measurements due to wind, gauge wetting, and gauge evaporation. The rescaled product better represents the extreme rainfall events at the coastline of Gulf of Alaska and Central America, while also preserving both the magnitude and spatial pattern in most other areas of North America.
- Bias in the NARR incoming shortwave radiation product, previously identified by comparing NARR estimates with observations at the Atmospheric Radiation Measurement (ARM) Southern Great Plains (SGP) site (Kennedy et al., 2010) and at 23 FLUXNET sites across North America (Wei et al., in prep.), was reduced by linearly rescaling the NARR values to match total daily downward shortwave radiation generated from the weather simulation model MTCLIM version 4.3. The rescaled NARR product better matches observed radiation at FLUXNET sites.
- Non-climate data sets, including land cover change, soil, phenology, and C3/C4 grass fractions, are not fully consistent with the land water mask maps derived from global and North American climate data. Grid cells indicated as water in these data sets but as land in land water mask were filled with values using a nearest neighbor method.
- The atmospheric CO2 concentration data prepared for the MsTMIP are consistent with the GLOBALVIEW-CO2 2011 data product (henceforth GV), the time series of historic atmospheric CO2 from Antarctic ice cores (MacFarling Meure et al., 2006), fossil fuel emissions (Marland et al., 2008), and atmospheric CO2 observations at Mauna Loa (MLO) and the South Pole (SPO). The MsTMIP atmospheric CO2 product before 1979, however, does not represent inter-annual variability other than that deriving from variability in fossil fuel emissions, and it does not include speculative changes in the magnitude or phase of annual cycles of CO2 in the atmosphere.
- The MsTMIP annual gridded nitrogen deposition data, compared to linear interpolation of Dentener (2006) nitrogen deposition data, capture an uneven increase trend across different regions, with East and South Asia, West Europe, and North America experiencing higher levels and faster increases of N deposition input.
- Different terrestrial biosphere models usually use different, and sometimes highly customized, schemes of land cover type parameterization and land cover data. For a carbon cycle model intercomparison project, like MsTMIP, with over 20 different participating models, this inconsistency would have added additional uncertainties and obstacles to the model comparison and diagnosis of model parameterizations. A consistent land cover data with a

land cover classification scheme suitable for carbon cycle modeling was an important criterion for the success of MsTMIP. The SYNMAP-Hurtt land use and land cover change product (referred to as SYNMAP-H) was chosen for MsTMIP based on its global coverage, inclusion of land use change fractions (required for a subset of participating models), overlap with the time horizon of MsTMIP simulations, and its use in the Intergovernmental Panel on Climate Change (IPCC) process. See Data Acquisition and Methods section for details.
- The merged land-use and land-cover change product does not exactly replicate the baseline SYNMAP product. This relates to the disparity in the extent of similar Hurtt and SYNMAP classes in 2000/2001. However, the overall aim of the harmonization was not to replicate the baseline SYNMAP per se. Rather, the goal was to translate the Hurtt product to a land cover classification that is more amenable to land surface modeling. Alternatively, MAST-DC regards Hurtt as the “best estimate” of land cover/use dynamics from 1700-2100 and they merely grafted onto this product the SYNMAP class structure.
- Soil bulk density values that are overestimated in Harmonized World Soil Database (HWSD) version 1.1 for Andosols and Histosols soil types were corrected using the corresponding depth-weighted average values from ISRIC-WISE, version 1.0 (Batjes, 2008). The correction mainly impacts the North American boreal region and a few places of southeastern Asia where Andosols and Histosols dominate.
- Comparison of a new Unified North American Soil Database (UNASD; Liu et al., 2013) prepared for MsTMIP to a subset of HWSD demonstrates a pronounced difference in the spatial distributions of soil properties and soil organic carbon mass. However, the UNASD provides more accurate and detailed information particularly in Alaska and central Canada.

5. Data Acquisition Materials and Methods:

Data Requirements

In order to meet the objectives of MsTMIP’s experimental design, the goal was to provide modeling teams, to the extent possible, with a complete and consistent set of environmental driver data. This required MAST-DC to make improvements in the quality and/or the spatial and temporal coverage and resolution of several of the original environmental data sets. In addition to being of high quality, the environmental driver input data chosen for MsTMIP also needed to meet the following requirements:

- Data sets must be compatible with over 20 different TBMs;
- Data sets must provide consistent spatial coverage for the land surface within the two simulation domains: (1) North American: 10 degrees to 84 degrees North latitude; -50 degrees to -170 degrees West longitude, and (2) Global: all land surface areas excluding Antarctica;
- Spatial resolutions must be compatible with the two sets of simulations: (1) North American (0.25 degrees by 0.25 degrees) and (2) Global (0.5 degrees by 0.5 degrees);
- Temporal resolution and extent must be compatible with the two sets of simulations: (1) North America (3-hourly, 1801-2010) and (2) Global (6-hourly, 1801-2010);
- Data sets must provide smooth transitions in time, without any unrealistic spikes or discontinuities; and
- Data sets must be physically consistent with one another. For example, climate, soil, and land use and land cover change history needed to represent the same land domain as indicated in the land-water mask, and the prescribed phenology data needed to be consistent with the time-varying land cover data for each time step.

For most data categories, the North American data sets are based on the same data sources as the global products. However, different climatology and soil data products were compiled for the two domains. This decision was driven primarily by the availability of these drivers at the spatial and temporal resolution needed for the regional simulations. In addition, by holding the source of other drivers constant between the global and North American simulations, MsTMIP created an opportunity to test the impact of the choice of climate and soil characteristics on model estimates.

Data Synthesis and Processing

Wei et al. (in review) describes how the MsTMIP model intercomparison driver data were compiled. This information is summarized below.

Global Climate: CRUNCEP. MAST-DC combined the Climate Research Unit CRU Time Series (TM) 3.2 (Mitchell and Jones, 2005) and the National Centers for Environmental Prediction (NCEP) / National Center for Atmospheric Research (NCAR) Reanalysis 1 (Kalnay et al., 1996) gridded global climatology data sets to produce the “CRUNCEP” global climate data set. The original data sets have the following characteristics:

- The CRU Time Series (TS) 3.2 monthly mean climatology time series data were constructed from meteorological stations covering the global land surface (excluding Antarctica) at a 0.5-degree resolution from 1901 to 2009. Nine climate variables, temperature, diurnal temperature range, daily minimum and maximum temperatures, precipitation, wet-day frequency, frost-day frequency, vapor pressure, and cloud cover, are provided in the CRU TS 3.2 data (Mitchell and Jones, 2005).
- The NCEP/NCAR) Reanalysis 1 surface data set provides 4-times daily surface or near the surface climatology values for 1948-01-01 to present at 2.5 degree x 2.5 degree (global grid) or 1.875 degree x 1.875-degree (T62 Gaussian grid) spatial resolution for the global. Eleven climate variables are provided, including air temperature, surface lifted index, best lifted index, omega, potential temperature, precipitable water, pressure, relative humidity, sea level pressure, U-wind, and V-wind (Kalnay et al., 1996).

The new CRUNCEP data set provides a globally gridded (0.5 degree by 0.5 degree) and sub-daily (6-hourly) time-varying climatology product that spans the period between 1901 and 2010. It contains seven climatology variables; including incoming longwave and shortwave radiations, pressure, air specific humidity, precipitation, temperature, and wind. In the process of creating this new climatology product, MAST-DC also corrected known biases in temperature and shortwave radiation in the NCEP/NCAR Reanalysis product described by Zhao et al. (2006). [Zhao et al. (2006) showed that NCEP/NCAR Reanalysis climatology overestimates downward shortwave radiation, especially in non-tropical regions, and underestimates surface temperature for almost all latitudes. Biases in climatological variables can introduce substantial errors into Gross Primary Productivity (GPP) and Net Primary Productivity (NPP) estimates.] By fusing NCEP/NCAR with the CRU climatology, MAST-DC forced the amplitude of CRUNCEP product to be consistent with the observation-based CRU climatology, while preserving the diurnal and daily variability in the NCEP/NCAR Reanalysis product.

Below is a summary of the fusion method used by MAST-DC:

- Between 1948 and 2009, the CRUNCEP data are based on CRU climatology. NCEP was used only to generate the diurnal and daily variability. The

NCEP was first bi-linearly interpolated to the 0.5-degree x 0.5-degree resolution of CRU for all variables except precipitation, where interpolation is based on the nearest neighbor. CRU provides cloud cover that was converted to incoming solar radiation based on calculation of clear sky incoming solar radiation as a function of date and latitude of each grid cell (Berliand, 1960). The relative humidity was converted to specific humidity as a function of temperature and surface pressure.

- The monthly average values, M, of each NCEP variable, were calculated on each 0.5 or .25-degree grid cell. Then the 6-hourly values were calculated as C_{6h}=C*m/M, except for temperature where C_{6h}=C+(m-M). C_{6h} is the 6-hourly CRUNCEP value to be calculated for each 0.5 degree grid cell, C is the monthly CRU value, and m is the NCEP 6-hourly value interpolated to CRUs 0.5-degree grid.
- After 2009, or the CRU ending period, MAST-DC extrapolated the fields directly from NCEP. This is done by doing a regression between the NCEP monthly mean values Mx(month,x,y) and the Monthly CRU averaged at the scale of NCEP (2.5 degree x 2.5 degree). For each grid cell, a median linear regression between NCEP and CRU for the period 1990 to 2009 was performed. Then C_{6h}= a*Mx+b, where a and b are the coefficients of the linear interpolation.
- Before 1948, the procedure is the same as for 1948-2009 except that for variability MAST-DC used NCEP data from 1948 and then the same intra-annual variability is repeated for every year.
- Climate variables: pressure, downward longwave radiation, and wind speed are not available in the CRU data set. MAST-DC directly used the variables coming from NCEP and interpolated them to the 0.5-degree x 0.5-degree CRU grid. Before 1948 MAST-DC took the data directly from 1948, hence there is no inter-annual variability for these variables before 1948.

North American Climate: NARR. The NCEP North America Regional Reanalysis (NARR) is a long-term, dynamically consistent, high-resolution, high-frequency, atmospheric and land surface hydrology data set (1979-present) for the North American domain (Mesinger et al., 2006). It has 3-hourly temporal resolution and 32 km spatial resolution. The NARR model uses the very high resolution NCEP Eta Model (32km / 45 layers) together with the Regional Data Assimilation System (RDAS) which, significantly, assimilates precipitation along with other variables. Seven mono-level NARR climate variables were selected for MsTMIP, including air temperature at 2 m (air.2m), accumulated total precipitation (apcp), downward longwave radiation flux (dlwrf), downward shortwave radiation flux (dswrf), relative humidity at 2 m (rhum.2m), specific humidity at 2m (shum.2m), and wind speed at 10 m (wnd.10m). The original NARR data were provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their web site at http://www.esrl.noaa.gov/psd/.

Additional processing was performed to convert original NARR data into the format required by MsTMIP and to further reduce the bias of certain NARR climate variables. This procedure included the regridding to 0.25 degree x 0.25 degree spatial resolution and improvements made to NARR precipitation and downward shortwave radiation flux.

The original NARR data are in Lambert Conformal Conic Projection and have a spatial resolution of 32 km, which is different from the Sphere-based Geographic Lat/Lon coordinate reference system and the 0.25 degree spatial resolution required by MsTMIP. The Spherical Coordinate Remapping and Interpolation Package (SCRIP) was utilized to regrid NARR variables to MsTMIP North America grids using two different methods: area-weighted average and distance-weighted average. Area-weighted average method was used for precipitation and radiation flux variables in order to preserve their total amounts in the North America region. Distance-weighted average method was used for other variables. The original NARR provides variables of U-direction wind speed (U-wnd.10m) and V-direction wind speed (V-wnd.10), which were combined together to calculate the surface wind speed variable prior to the regridding process.

Sun and Barros (2010) compared NARR with National Climatic Data Center (NCDC) rain gauge measurements and found that although NARR reproduces the spatial patterns of NCDC parameters, the frequency of large rainfall events, the magnitude of maximum rainfall, and the mean intermittency are underestimated. Xie et al. (2003) found that the 2.5° pentad Global Precipitation Climatology Project (GPCP) analyses, which matched the monthly GPCP in magnitude, reproduced spatial distribution patterns of total precipitation with relatively high quality especially over land. Thus MAST-DC rescaled NARR 3-hourly precipitation with the GPCP monthly gridded data set version 2.1. The GPCP v2.1 consists of monthly total precipitation derived from satellite and gauge measurements (Adler et al., 2003). Although the GPCP product has a relatively coarse resolution of 2.5 degree, it presented the advantage of including a correction to compensate for systematic biases in gauge measurements due to wind, gauge wetting, and gauge evaporation. Since the NARR data set does not include such corrections, the complimentary use of the GPCP product helped to identify the effect of systematic errors in precipitation measurements. For each month, precipitation of all 3-hourly 0.25 degree NARR grids within each 2.5 degree GPCP grid were summed up and adjusted linearly to match the total precipitation amount in GPCP. MAST-DC analysis showed that rescaled NARR precipitation preserved both the magnitude and spatial pattern in most area of North America. Precipitation was reduced after rescaling in a few areas, including lower Alaska and Montana. Extreme rainfall events at the coastline of Gulf of Alaska and Central America were enhanced by the rescaling process.

Kennedy, et al. (2010) found NARR has significant positive bias for incoming shortwave radiation flux (dswrf) under clear- and all-sky conditions compared with ARM Southern Great Plains (SGP) site continuous forcing during the period 1999-2001. Comparing NARR incoming shortwave radiation flux with FLUXNET tower observations showed that NARR downward shortwave radiation is overestimated by about 30% overall with higher positive bias under cloudy conditions. The MTCLIM model version 4.3, a weather simulation model, was used to reduce this shortwave radiation bias. Given input data from one location, MTCLIM generates weather information for another location with potentially different elevation, slope, and aspect from the input location (Running et al., 1987, Thornton and Running, 1999). The input location is referred to as the “base station” while the new location for output is referred to as the “site.” In the case of NARR, the “site” is same as the “base station” which is the center point of every 0.25 degree grid cell in North America. The MTCLIM model was fed daily max/min temperature and daily total precipitation which were derived from the 3-hourly NARR original temperature and rescaled precipitation for each grid. The model then calculated the daily total shortwave radiation flux for that grid. The 3-hourly NARR dswrf variable values were then adjusted to match the total daily downward shortwave radiation generated from MTCLIMl. The rescaled NARR downward shortwave radiation have a 25%-30% decrease compared with original NARR data. Between 70°N and 80°N, the original NARR downward shortwave radiation almost has a flat curve. This issue was addressed in the reanalyzed NARR data.

Land-Water Mask. The MsTMIP land-water mask maps contain values 0 and 1, with 0 indicating water and 1 indicating land. The mask maps specify all land grid cells on which MsTMIP global and regional simulations were run. The selection criterion for land-water mask maps was to be consistent with climate driver data. For global land-water mask map, MAST-DC took the CRU-NCEP land water mask directly. For North America land-water mask map, MAST-DC took the original NARR mask, regridded it to 0.25 degree spatial resolution to derive a land fraction map using an area-weighted average method to preserve the total amount of land area, then used a threshold of 50% to classify regridded land fraction map into land-water mask map.

Atmospheric CO2 Concentration. The atmospheric CO2 concentration data prepared for MsTMIP is consistent with the GLOBALVIEW-CO2 2011 (henceforth GV) data product, the time series of historic atmospheric CO2 from Antarctic ice cores (MacFarling Meure et al., 2006), fossil fuel emissions (Marland et al., 2008), and Scripps CO2 program (SIO) atmospheric CO2 observations at Mauna Loa (MLO) and the South Pole (SPO). During the period of 1979-2010, when direct observations are available, CO2 concentrations were set directly to the GV marine boundary reference surface, interpolated from GV's native latitude-time grid to that needed for MsTMIP simulations. For the period prior to 1979, MAST-DC preserved the mean annual cycle from GV and imposed this on a modeled CO2 surface that represented annual mean concentrations and a time-evolving meridional gradient. Following the methods of Conway and Tans (1999), the annual mean difference between Mauna Loa and South Pole in the GV product was modeled as a linear function of fossil fuel (FF) emissions (Marland et al., 2008). Extrapolated to zero FF emissions, the preindustrial MLO-SPO difference estimated in this manner is 0.3 ppm. Performing this same exercise using SIO observations instead of GV yielded a stronger dependence of the meridional gradient on FF emissions and a preindustrial MLO-SPO difference of -1.2 ppm. While it is possible that preindustrial southern hemisphere CO2 values exceeded those in the northern hemisphere (Conway and Tans, 1999), MAST-DC judged that it was more parsimonious to assume a small preindustrial inter-hemispheric CO2 gradient that the GV-based scheme achieves natively. The MsTMIP product agrees well with Scripps CO2 data before 1979 at SPO and MLO, and with Law Dome ice core data in Antarctica data (MacFarling Meure et al., 2006). It does not represent inter-annual variability other than that derived from variability in FF emissions, and it does not include speculative changes in the magnitude or phase of annual cycles of CO2 in the atmosphere.

Nitrogen Deposition. MAST-DC used the approach described in Tian et al. (2010) and Lu et al. (2012) to create a time-varying annual nitrogen deposition data set for both global (0.5 degree by 0.5 degree resolution) and North American (0.25 degree by 0.25 degree resolution) simulations based on Dentener (2006) maps of total inorganic nitrogen (N), NHx (NH3 and NH4+), and NOy (all oxidized forms of nitrogen other than N2O) deposition and to introduce spatial and temporal variations from nitrogen emissions. The original maps cover the years 1860, 1993, and 2050 and have a spatial resolution of 5 degree longitude by 3.75 degree latitude.

For the Dentener (2006) maps, the annual variation of nitrogen deposition rate from year 1890 to year 1990 was controlled by EDGAR-HYDE 1.3 nitrogen emission data (Van Aardenne et al., 2001) which provides information on annual totals of NH3 and NOx emissions from 10 anthropogenic sources within 1.0 ×1.0-degree grid cells for each decade. MAST-DC assumed that the temporal trends of NHx-N and NOy-N depositions were consistent with those of NH3 and NOx emissions between 1890 and 1990. MAST-DC also assumed that nitrogen deposition increased linearly over the time periods 1860-1890 and 1990-2050. Following these assumptions and the methods of Tian et al. (2010), the annual global 0.5 degree and North America 0.25 degree nitrogen deposition estimates were calculated by temporally and spatially interpolating the original Dentener's nitrogen deposition data.

Specifically, the development of time-series nitrogen deposition data during 1890-1990 followed several assumptions:

- Nitrogen deposition varied along with N emissions:

$$ND_i = ND_{i-1} + (E_{Dj} - E_{Dj-1}) / 10 \times (ND_{1990} - ND_{1890}) / (E_{1990} - E_{1890})$$

Where ND_i is N deposition rate for NH_x or NO_y in a specific year i (unit: mg $N/m^2/yr$); correspondingly, ED_j and ED_{j-1} are emission rates of NH_3 or NO_x in decade j and j-1.

- In a few grid cells (< 0.1%), where change trend of nitrous gas emissions was inconsistent with nitrogen deposition, e.g., emission rate in 1990 is less than that in 1890, while nitrogen deposition showed the opposite pattern, MAST-DC used linear change trend to retrieve the annual variation of nitrogen deposition. If emission change between two consecutive decades was 3 times larger than that between 1890 and 1990 (i.e., abrupt change occurred, in very few cells, < 0.01%), MAST-DC used linear change rate, instead of emission-based temporal variation.

- For N deposition in ocean areas, MAST-DC used linear trend to retrieve the annual pattern because N emission data points were mainly located on land.

Biome. The static satellite-based land cover product SYNMAP (Jung et al., 2006) was chosen for MsTMIP biome classification due to its: (1) reconciliation of multiple global land cover products (i.e., Global Land Cover Characterization Database (GLCC) (Hansen et al., 2000; Loveland et al., 2000), GLC2000 (2003), and the 2001 MODIS land cover product (Friedl et al., 2002); (2) global coverage at 1-km resolution; (3) general definition of classes based on life form, leaf type, and leaf longevity which allowed for simple mapping rules to plant functional types (PFTs) used in different TBMs, and (4) assumed representation of year 2000/2001 land cover/biome status. Generality was a key concern as PFT schemes used in TBMs vary widely. The original 1-km resolution SYNMAP data was upscaled by MAST-DC to 0.5-degree resolution for global and 0.25-degree for North America to derive two variables: land cover class fraction and dominant land cover class. The land cover class fraction variable provides fraction of each of the 48 land cover classes contained in each grid cell and the class with highest fraction value was selected as the dominant land cover class for that grid cell.

Land-use and Land-cover Change. For this data set, land-use and land-cover change was prescribed by merging SYNMAP (Jung et al., 2006) with the time-varying land use harmonization data from the fifth Assessment Report (AR5) of the Intergovernmental Panel on Climate Change (IPCC) (Hurtt et al., 2011). The SYNMAP-Hurtt product was chosen for MsTMIP based on its global coverage, inclusion of land use change fractions (required for a subset of participating models), overlap with the time horizon of MsTMIP simulations, and its use in the IPCC process.

The land use harmonization product (Hurtt et al., 2011) provides mapped fractional coverage and underlying annual land use transitions for six land use classes (primary land, secondary land, cropland, pasture, urban, and barren) at 0.5-degree by 0.5-degree spatial resolution. Inputs include new gridded historical maps of crop and pasture data from HYDE 3.1 for 1500–2005, updated estimates of historical national wood harvest and of shifting cultivation, and future information on crop, pasture, and wood harvest from the Integrated Assessment Model (IAM) implementations of the Representative Concentration Pathways (RCPs) for the period 2005–2100. The historical land use harmonization data (1801-2005) were combined with the Representative Concentration Pathway (RCP) 4.5 scenario (2006-2010) to match the time horizon of MsTMIP model simulations (1801-2010).

SYNMAP and Hurtt were merged using one-to-one (direct mapping) and one-to-many mapping rules based on map intersection during their period of overlap, i.e., both products exist for 2000/2001. Direct mappings were used for those Hurtt types with direct analogues in the SYNMAP class structure (i.e., the class denoting urban and built-up land). For the remaining five Hurtt types, MAST-DC used a one-to-many mapping approach based on the temporal overlap of the two products in 2000/2001.

Table 16. One-to-many mapping rules used to merge Hurtt with SYNMAP

Hurtt Class	SYNMAP Target Class
Croplands	Trees & Crops[1], Shrubs & Crops, Grasses & Crops, Crops[2]
Pasture	Trees & Grasses[1], Shrubs, Shrubs & Grasses, Shrubs & Crops, Grasses, Grasses & Crops
Primary	Trees[1], Shrubs, Shrubs & Grasses, Shrubs & Crops, Shrubs & Barren
Secondary	Trees1, Trees & Shrubs[1], Trees & Grasses1, Trees & Crops1, Shrubs, Shrubs & Grasses, Shrubs & Crops
Barren	Shrubs & Barren, Grasses & Barren, Barren, Snow & Ice, Water[3]

Notes:

[1]All tree leaf types and longevities.

[2]The pure SYNMAP crop class was favored in the one-to-many mappings (relative weight increased by a factor of 5). This was based on Hurtt croplands being pure with in-pixel admixtures of other classes accounted for by Hurtt non-croplands as opposed to SYNMAPs mixed classes.

[3]The land mask used was all grid cells where the SYNMAP water class was less than unity.

One-to-many mappings involved three steps. Using an example of mapping Hurtt pasture to SYNMAP shrubs and grasses, the following processing took place:

- **Perform baseline map intersection**. Hurtt and SYNMAP were intersected grid cell by grid cell during 2000/2001. For the Hurtt product MAST-DC used the average fractional coverage across 2000/2001 whereas the SYNMAP baseline product referenced 2000/2001 natively. Map intersection was accomplished by tabulating the SYNMAP fractional coverages for allowable target classes only and matching Hurtt fractional coverage. For example: Assume that a given intersected grid cell shows Hurtt pasture with a fractional coverage of 0.5. In that same grid cell the SYNMAP product has only two eligible target classes: shrubs and grasses. Both of these have fractional coverages of 0.25. Note that while other SYNMAP classes may exist in the pixel (e.g., any forested type), these are not mapped to pasture.
- **Normalize the fractional coverages**. The tabulation of fractional coverages from Step 1 was then normalized to unity. Continuing the example from above, it follows that Hurtt pasture in this pixel is composed equally of the SYNMAP shrubs and grasses classes. That is, one unit of Hurtt pasture equals a half-unit each of SYNMAP shrubs and SYNMAP grasses.
- **Apply the normalized fractional coverages in time**. The normalized fractional coverages were then applied to the same pixel at all time steps. For this example, take that same pixel in 1940, now with a fractional coverage of 0.4 for Hurtt pasture. Based on Step 2 MAST-DC used the one-to-many mapping and normalized fractional coverages to apportion 0.4 Hurtt pasture to 0.2 each of the SYNMAP classes shrubs and grasses.

Applying these rules resulted in the merged dynamic land cover product from 1700-2100 with all 48 SYNMAP classes on an annual time step.

C3 and C4 Grass Fractions. MAST-DC created global 0.5 degree C3 and C4 grasses relative fraction maps under the “present” climate state based on the CRU-NCEP mean monthly precipitation and temperature data between 2000-2010 using an approach described in Still et al. (2003) based on growing season temperature. For grid cells characterized as grasslands (or containing grasslands) the relative fraction map defines the fraction of those grasses that are C3 or C4, so that in each grid cell the C3 and C4 grass fractions sum to one regardless of the total percentage of grassland contained in the grid cell. The value is zero if no grass is present in a particular grid cell.

The grass fractions were calculated as follows:

Among the 12 months, if there was at least one month with monthly mean air temperature (T) above 22 degrees C and at the same time the monthly total rainfall (P) was above 25 mm in a grid cell, it was assumed that the C4 grass relative fraction to be equal to the number of months where C4 photosynthesis is favored relative to the number of growing season months with air T greater than 5 degrees C. Therefore, C4 grass relative fraction was calculated as:

[C4 grass relative fraction] = [number of months with T>22 degrees C and P>25 mm] / [number of months with T>5 degrees C]

[C3 grass relative fraction] = 1 − [C4 grass relative fraction]

The C3 and C4 grass fraction maps were then created through merging the C3 and C4 relative fraction maps with SYNMAP:

[C3/C4 grass fraction] = [C3/C4 grass relative fraction] − [SYNMAP grass fraction]

SYNMAP contains 13 land cover classes that include grasses land cover class, with 12 of them mixtures of grasses with trees/shrubs/crops/barren. For the mixed classes, it was assumed that grasses accounted for 50% area of a cell. The SYNMAP grass fraction in each cell was calculated as the sum of the grass fraction of all different classes included in the cell.

The North American C3 and C4 relative grassland fraction maps were created using the same approach with NARR climate data.

MsTMIP only provides a constant C3/C4 data product under "present" climate conditions. For models that needed time-varying C3/C4 grass fractions, the same approach was applied to historical land cover data and historical precipitation / temperature climate data to generate C3/C4 grassland maps for previous years.

Major Crops.MAST-DC identified and extracted four globally significant crop types (maize, rice, soybean, and wheat) from the Monfreda et al. (2008) global crop database for 2000 at a 5 min by 5 min (approximately 10 km by 10 km) spatial resolution. The original data was resampled to 0.5 degrees by 0.5 degrees (global) and 0.25 degrees by 0.25 degrees (North American) spatial resolutions. These major crop designations do not provide detailed model simulation prescription, but rather guidance for models that need to specify crop types or cropping systems.

Phenology. For models that use remote sensing products to prescribe plant phenology to calculate GPP or NPP, MAST-DC constructed monthly maps of Normalized Difference Vegetation Index (NDVI), Leaf Area Index (LAI), and absorbed fraction of Photosynthetically Active radiation (fPAR) consistent with the harmonized Hurtt-SYNMAP land cover change data on both global and North American grids for 1801-2010.

LAI and fPAR values were derived from the Global Inventory Monitoring and Modeling System version g (GIMMSg) NDVI data set (Tucker et al., 2005) which is a 15-day maximum value composite calculated from AVHRR at about 8km spatial resolution for 1982-2010 and adjusted for missing data, satellite orbit drift, sensor degradation, and volcanic aerosols. The 15-day GIMMSg NDVI for 1982-2010 was first converted to monthly mean composites for 12 months from January to December. This averaging process reduced noise in the data such as sudden and large changes due to cloud contamination. The monthly mean composites were then regridded to the 0.5 degrees grid for global and the 0.25 degrees grid for North America. MAST-DC then calculated the average seasonal cycle in monthly mean NDVI values and used them to calculate LAI and fPAR using methods described in Schaefer et al. (2002).

To harmonize phenology data with the the harmonized Hurtt-SYNMAP land cover change data, MAST-DC assumed that a pixel would consist of tiles, each corresponding to a different land use/cover class with fractional areas set by the MsTMIP coverage maps as a function of year from 1801 to 2010. Maps of LAI and fPAR were calculated assuming the entire land surface was one of the 12 SiB biome classes (Sellers et al., 1986) resulting in 12 sets of LAI and fPAR maps corresponding to the 12 SiB biome classes, all calculated from the same NDVI values but using different parameter values unique to each biome (Sellers et al., 1996). The 12 SiB biomes were then mapped to the 47 SYNMAP land use/cover types using one-to-one or one-to-many mapping, resulting in 47 sets of LAI and fPAR maps corresponding to the 47 SYNMAP classes. This two-step process was required because the parameters used to calculate LAI and fPAR were not available for each of the 47 SYNMAP types. Combining these 47 sets of LAI and fPAR maps and the yearly MsTMIP land-use and land-cover change data, time-evolving and land use/cover type explicit LAI and fPAR data products were created. If a grid cell did not contain a particular SYNMAP type for a specific year, a standard missing value was inserted into the corresponding LAI and fPAR maps. Participating model would then extract the LAI and fPAR values for a particular SYNMAP class and use them for the corresponding tile.

Global Soil: Gridded HWSD. Harmonized World Soil Database (HWSD) version 1.1 (FAO/IIASA/ISRIC/ISS-CAS/JRC, 2011) was used as the source for MsTMIP global soil data. Each soil mapping unit in the HWSD is composed of several different soil units (or soil types) defined by major soil group code following a combined FAO-74/FAO-85/FAO-90 soil classification system. For the global simulations, the original HWSD was regridded to a spatial resolution of 0.5 degrees by 0.5 degrees by selecting the dominant soil type within each grid cell. Eight (8) physical and chemical soil properties associated with the dominant soil type in each soil layer were then selected (Table 14 above). One additional property, reference soil depth, was extracted from HWSD and provided as a proxy for mineral soil depth, even though this reference soil depth is not precise. Bulk density values that are overestimated in HWSD v1.1 for Andosols and Histosols soil types were corrected using the corresponding depth-weighted average values from ISRIC-WISE, version 1.0 (Batjes, 2008). The correction mainly impacts the North American boreal region and a few places of southeastern Asia where Andosols and Histosols dominate.

North American Soil: Unified North American Soil Database (UNASD). A new gridded database of harmonized soil physical and chemical properties for North America was created for MsTMIP by fusing the most recent regional soil information from U.S. STATSGO2, Canada soil databases versions 3.2 and 2.2, and the HWSD v1.1. The fused database was then harmonized into two standardized soil layers as for the HWSD. The top soil layer ranges from 0 to 30 cm and the sub soil layer ranges from 30 to 100 cm. A comparison with the subset of HWSD demonstrated pronounced difference in the spatial distributions of soil properties and soil organic carbon mass between the UNASD and HWSD, but overall the UNASD provides more accurate and detailed information particularly in Alaska and central Canada. The methods used to develop the UNASD and the comparisons with HWSD are described in detail in Liu et al. (2013).

6. Data Access:

This data set is available through the Oak Ridge National Laboratory (ORNL) Distributed Active Archive Center (DAAC).

Data Archive Center:

Contact for Data Center Access Information:
E-mail: uso@daac.ornl.gov
Telephone: +1 (865) 241-3952

7. References:

Adler, R.F., Huffman, G.J., Chang, A., Ferraro, R., Xie, P.P., Janowiak, J., et al. 2003. The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979-present). J. Hydrometeorol. 4: 1147-1167. DOI:10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2

Batjes, N.H. 2008. ISRIC-WISE harmonized global soil profile dataset (Version 3.1). ISRIC-World Soil Information, Wageningen.

Berliand, T.G. 1960. Method of climatological calculation of global radiation. Meteorologiya i Gidrologiya 6: 9-12.

Conway, T.J., and P. P. Tans. 1999. Development of the CO2 latitude gradient in recent decades. Global Biogeochem. Cycles 13(4): 821-826. DOI:10.1029/1999GB900045

Dentener, F.J. 2006. Global maps of atmospheric nitrogen deposition, 1860, 1993, and 2050. Data set. Available on-line [http://daac.ornl.gov/] from Oak

Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A. DOI:10.3334/ORNLDAAC/830

Denning, A.S., et al. 2005. Science implementation strategy for the North American Carbon Program: A Report of the NACP Implementation Strategy Group of the U.S. Carbon Cycle Interagency Working Group. U.S. Carbon Cycle Science Program, Washington, DC. 68 pp.

FAO/IIASA/ISRIC/ISS-CAS/JRC. 2011. Harmonized World Soil Database (version 1.1). FAO, Rome, Italy and IIASA, Laxenburg, Austri.

Friedl, M.A., McIver, D.K., Hodges, J.C.F., Zhang, X.Y., Muchoney, D., Strahler, A.H., Woodcock, C E., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F., and Schaaf, C. 2002. Global land cover mapping from MODIS: algorithms and early results. Remote Sens. Environ. 83: 287-302. DOI:10.1016/S0034-4257(02)00078-0

Galloway, J.N., Dentener, F.J., Capone, D.G., Boyer, E.W., Howarth, R.W., Seitzinger, S.P., Asner, G.P., Cleveland, C., Green, P., Holland, E., Karl, D.M., Michaels, A.F., Porter, J.H., Townsend, A., and Vorosmarty, C. 2004. Nitrogen cycles: past, present, and future. Biogeochemistry 70: 153-226. DOI:10.1007/s10533-004-0370-0

Global Land Cover 2000 database (GLC2000). 2003. European Commission, Joint Research Centre. [http://bioval.jrc.ec.europa.eu/products/glc2000/glc2000.php]

Hansen, M.C., Defries, R.S., Townshend, J.R.G., and Sohlberg, R. 2000. Global land cover classification at 1km spatial resolution using a classification tree approach. Int. J. Remote Sens. 21: 1331-1364. DOI:10.1080/014311600210209

Huntzinger, D.N., Schwalm, C., Michalak, A.M., King, A.W., Schaefer, K., Jacobson, A R., Wei, Y., Liu, S., Cook, R B., and Post, W.M. 2013. The North American Carbon Program Multi-scale synthesis and Terrestrial Model Intercomparison Project -- Part I: Overview and experimental design. Geosci. Model Dev. Discuss. 6: 3977-4008. DOI:10.5194/gmdd-6-3977-2013

Hurtt, G.C., Chini, L., Frolking, S., Betts, R., Edmonds, J., Feddema, J., Fisher, G., Goldewijk, K.K., Hibbard, K., Houghton, R., Janetos, A., Jones, C., Kinderman, G., Konoshita, T., Riahi, K., Shevliakova, E., Smith, S.J., Stefest, E., Thomson, A.M., Thornton, P., van Vuuren, D., and Wang, Y. 2011. Harmonization of land-use scenarios for the period 1500-2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. Climatic Change 109: 117-161. DOI:10.1007/s10584-011-0153-2

Jung, M., Henkel, K., Herold, M., and Churkina, G. 2006. Exploiting synergies of global land cover products for carbon cycle modeling. Remote Sens. Environ. 101: 534-553. DOI:10.1016/j.rse.2006.01.020

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., et al. 1996. The NCEP/NCAR 40-year reanalysis project. Bulletin of the American Meteorological Society 77(3): 437-471. DOI:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2

Kennedy, A., Dong, X., Xi, B., Xie, S., Zhang, Y., and Chen, J. 2010. A comparison of MERRA and NARR reanalysis with the DOE ARM SGP continuous forcing data. AGU fall meeting 2010, San Francisco, California, USA, 13-17 December, 2010.

Liu, S., Wei, Y., Post, W.M., Cook, R.B., Schaefer, K., and Thornton, M.M. 2013. The Unified North American Soil Map and its implication on the soil organic carbon stock in North America. Biogeosciences 10: 2915-2930. DOI:10.5194/bg-10-2915-2013

Loveland, T.R., Reed, B.C., Brown, J.F., Ohlen, D.O., Zhu, Z., Yang, L., et al. 2000. Development of a global land cover characteristics database and IGBP DISCover from 1km AVHRR data. Int. J. Remote Sens. 21: 1303-1330. DOI:10.1080/014311600210191

Lu, C., Tian, H., Liu, M., Ren, W., Xu, X., Chen, G., and Zhang, C. 2011. Effects of nitrogen deposition on China's terrestrial carbon uptake in the context of multiple environmental changes. Ecol. Appl. 22: 53-75 DOI:10.1890/10-1685.1

MacFarling Meure, C., Etheridge, D., Trudinger, C., Steele, P., Langenfelds, R., van Ommen, T., Smith, A., and Elkins, J. 2006. Law Dome CO2 , CH_4 and N_2O ice core records extended to 2000 years BP. Geophys. Res. Lett. 33: L14810. DOI:10.1029/2006GL026152

Marland, G., Boden, T.A., and Andres, R.J. 2008. Global, Regional, and National Fossil Fuel CO2 Emissions. In Trends: A Compendium of Data on Global Change. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A. [http://cdiac.ornl.gov/trends/emis/overview]

Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P.C., Ebisuzaki, W., et al. 2006. North American regional reanalysis. Bull. Amer. Meteor. Soc. 87: 343-360. DOI:10.1175/BAMS-87-3-343

Mitchell, T.D., and Jones, P.D. 2005. An improved method of constructing a database of monthly climate observations and associated high-resolution grids, Int. J. Climatol. 25: 693-712. DOI:10.1002/joc.1181

Monfreda, C., Ramankutty, N., and Foley, J. A. 2008. Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. Global Biogeochem. CY 22: GB1022. DOI:10.1029/2007GB002947

Running, S.W., Nemani, R.R., and Hungerford, R.D. 1987. Extrapolation of synoptic meteorological data in mountainous terrain and its use for simulating forest evaporation and photosynthesis. Can. J. Forest. Res. 17: 472-483. DOI:10.1139/x87-081

Schaefer, K., Denning, A.S., Suits, N., Kaduk, J., Baker, I., Los, S., and Prihodko, L. 2002. Effect of climate on interannual variability of terrestrial CO2 fluxes. Global Biogeochem. CY 16(4), 49-1 - 49-12. DOI:10.1029/2002GB001928

Sellers, P.J., Mintz, Y., Sud, Y.C., and Dalcher, A. 1986. A simple biosphere model (SiB) for use within general circulation model. J. Atmos. Sc 43(6): 505-531. DOI:10.1175/1520-0469(1986)043<0505:ASBMFU>2.0.CO;2

Sellers, P.J., Los, S.O., Tucker, C.J., Justice, C.O., Dazlich, D.A., Collatz, G.J., and Randall, D.A. 1996. A revised land surface parameterization (SiB2) for Atmosphertic GCMs, Part II: The generation of global fields of terrestrial biophysical parameters from satellite data. J. Climate, 9(4): 706-737.

DOI:10.1175/1520-0442(1996)009<0706:ARLSPF>2.0.CO;2

Still, C.J., Berry, J.A., Collatz, G.J., and DeFries, R.S. 2003. Global distribution of C3 and C4 vegetation: Carbon cycle implications. Global Biogeochem. CY 17(1): 1-14. DOI:10.1029/2001GB001807

Sun, X. and Barros, A.P. 2010. An evaluation of the statistics of rainfall extremes in rain gauge observations, and satellite-based and reanalysis products using universal multifractals. J. Hydrometeorol. 11: 388-404. DOI:10.1175/2009JHM1142.1

Thornton, P.E., and S.W. Running. 1999. An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. Ag. For. Met. 93: 211-228. DOI:10.1016/S0168-1923(98) 00126-9

Tian, H. Q., Xu, X., Liu, M., Ren, W., Zhang, C., Chen, G., and Lu, C. 2010. Spatial and temporal patterns of CH_4 and N_2O fluxes in terrestrial ecosystems of North America during 1979-2008: application of a global biogeochemistry model. Biogeosciences 7: 2673-2694. DOI:10.5194/bg-7-2673-2010

Tucker, C.J., Pinzon, J.E., Brown, M.E., Slayback, D.A., Pak, E.W., et al. 2005. An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. Int. J. Remote Sens. 26: 4485-4498. DOI:10.1080/01431160500168686

van Aardenne, J. A., Dentener, F. J., Olivier, J. G. J., Klein Goldewijk, C. G. M., and Lelieveld, J. 2001. A 1 x 1 degree resolution dataset of historical anthropogenic trace gas emissions for the period 1890-1990. Global Biogeochem. CY 15(4): 909-928. DOI:10.1029/2000GB001265

Wei, Y., Liu, S., Huntzinger, D. N., Michalak, A. M., Viovy, N., Post, W. M., Schwalm, C. R., Schaefer, K., Jacobson, A. R., Lu, C., Tian, H., Ricciuto, D. M., Cook, R. B., Mao, J., and Shi, X.:. in review. The North American Carbon Program Multi-scale Synthesis and Terrestrial Model Intercomparison Project – Part 2: Environmental driver data, Geosci. Model Dev. Discuss., 6, 5375-5422, doi:10.5194/gmdd-6-5375-2013, 2013.

Wofsy, S.C., and R.C. Harriss. 2002. The North American Carbon Program (NACP). Report of the NACP Committee of the U.S. Interagency Carbon Cycle Science Program. U.S. Global Change Research Program, Washington, DC. 56 pp.

Xie, P., Janowiak, J.E., Arkin, P.A., Adler, R., Gruber, A., Ferraro, R., Huffman, G.J., and Curtis, S. 2003. GPCP pentad precipitation analyses: An experimental dataset based on gauge observations and satellite estimates. J. Climate 16: 2197-2214. DOI:10.1175/2769.1

Zhao, M., Running, S.W., and Nemani, R.R. 2006. Sensitivity of Moderate Resolution Imaging Spectroradiometer (MODIS) terrestrial primary production to the accuracy of meteorological reanalyses. J. Geophys. Res-Biogeo. 111: G01002. DOI:10.1029/2004JG000004