PalEON Model-Data Inter-comparison Project
Phase 2: Regional Runs

Overview

Summary:

This document describes the protocol for running the regional PalEON model intercomparison project (MIP 2, formerly Phase 1b) over the full PalEON spatial and temporal domains. This phase follows the site-level runs completed in summer 2015. While it is not required that all models have completed site-level runs in the past, it is highly encouraged that any new models complete these runs for diagnostic and trouble-shooting purposes.

Initial regional modeling efforts should be completed by **12 November 2015** so that any concerns or adjustments that need to be made prior to completion of the final runs can be discussed at the PalEON Berkeley Annual Meeting prior to AGU (12-13 December 2015). Upon completion of Phase 2 regional runs in early 2016, Phase 3 will commence, which is focused on active data assimilation into terrestrial ecosystem models.

Phase 2 Goals:

1. Perform ecosystem simulations over a wider range of environmental and meteorological conditions than found in the site-level runs from the PalEON MIP Phase 1 (formerly Phase 1a)
2. Compare modeled terrestrial ecosystem dynamics with spatially-explicit empirical and statistical PalEON data products such as the STEPPS pollen and Settlement Vegetation data products

Timeline/Key Dates

- **Initial Protocol Feedback Deadline –** 1 October 2015
- **Final Protocol Distributed & Met Drivers Posted on iPlant –** 12 October 2015
- **Preliminary Runs & Feedback –** 12 November 2015
- **Presentation & Discussion of Preliminary Results –** Berkeley Annual Meeting (12-13 December 2015)
- **Final Run Submission –** 29 February 2016 (tentative)

Meteorological & Environmental Driver Datasets

Summary of drivers:

Modeling teams are asked to run simulations regionally over the PalEON domain at the **0.5°x0.5° spatial** resolution of the provided meteorology drivers. This corresponds to an 80 x 30 grid representing longitudes **-100° to -60° (west)**, and latitude **35° to 50° (north)**. These drivers are in 0.5˚ grids and are available in three temporal resolutions:

1) sub-daily, 6-hourly temporal resolution in GMT where each value represents the mean of the subsequent 6 hours; 2) daily means; 3) monthly means.

Meteorology drivers were derived from the CCSM4 fully coupled Last Millennium (p1000, 850-1849) and AR-5 historical (1850-1900) PMIP3 inter-comparison and CRUNCEP data product (1901-2010). Details on the bias correction process to align the two CCSM4 products with CRUNCEP as well as artifacts and biases found in the PalEON met drivers can be found in a document describing Phase 1 drivers. Additional figures and animations illustrating the spatial and temporal patterns of the met drivers over the PalEON modeling domain can be found in the "met_qaqc" in the folder with the raw, sub-daily meteorology drivers.

Downloading Instructions:

All simulations should be run using the standardized driver data, which can be found in the iPlant Discovery Environment:

https://de.iplantcollaborative.org/de/?type=data&folder=/iplant/home/crollinson/paleon

To gain access to the iPlant repository, you will need to register and then contact one of the modeling personnel listed at the end of this document so you can be given access to the PalEON group modeling repository.

Meteorological driver data can be found in the "phase2_met_drivers" folder. **The most recent bias-corrected data as of 1 October 2015 is phase2_met_drivers_v2.** Inside of this folder there are drivers available in three temporal resolutions:

1) subdaily: phase2_met_regional_v2 (contains met_qaqc)
    a. contains folders with drivers compressed (.tar.bz2) in 100-year chunks
    b. driver format: 1 file per month
2) daily: phase2_met_regional_v2_daily
    a. contains a single folder (met_zip) with a single .tar.bz2 file per met driver variable
    b. driver format: 1 file per month
3) monthly: phase2_met_regional_v2_monthly.tar.bz2
    a. this is a single compressed file with all met drivers
    b. driver format: 1 file for entire temporal domain (850-2010)

Meteorological data can be downloaded from iPlant by checking that file and then selecting "simple download" in the "Download" menu in the discovery environment. iPlant also offers iDrop, a desktop app that can be used to download or upload data, and iCommands, a set of command line tools, both of which tend to be easier to use and quicker than using the web interface.

Meteorological Driver Descriptions:

ALMA Name (CF equiv.)	ALMA Definition	Units	Prec.
lat (lat)	latitude	degrees_ north	float
lon (lon)	longitude	degrees_east	float
time (time)	time	Calendar date (Gregorian proleptic)	doubl e
lwdown [time,lat,lon] (fldlw)	Incident longwave radiation averaged over the time step of the forcing data	W m-2	float
swdown [time,lat,lon] (fldsw)	Incident radiation in the shortwave part of the spectrum averaged over the time step of the forcing data	W m-2	float
precipf [time,lat,lon] (rainfall_flux + snowfall_flux)	The per unit area and time precipitation representing the sum of convective rainfall, stratiform rainfall, and snowfall	kg m-2 s-1	float
psurf [time,lat,lon] (ps)	Pressure at the surface	Pa	float
qair [time,lat,lon] (q)	Specific humidity measured at the lowest level of the atmosphere	kg kg-1	float
tair [time,lat,lon] (air_temperature)	2 meter near surface air temperature	K	float
wind [time,lat,lon]	Wind speed measured with a vertical coordinate in height of 10 m	m s-1	float

Environmental Driver Datasets

Data Type	**Source**	**Temporal Extent**	**Temporal Resolutio n**	**Details**
Domain Mask	PalEON met data	Static (One Time)	Static (One Time)	This is the land mask to be used for these runs. All environmental drivers have been clipped & resampled to match this mask. Use this to resample and mask any additional variables your model uses.
CO2	PMIP3 + NOAA	850-2010	Annual, Monthly	Developed by splicing the PMIP-3 Law Dome composite time series (0-2000), and NOAA's Mauna Loa baseline CO_2 observatory data (2001-2010) with monthly variability from MsTMIP added in

				post-hoc (monthly data only)
Land-Use	Chini et al. Harmonized Global Land Use for Years 1500-2010	850-2010	Annual	These data are not to be confused with the Hurtt (2006) data. These were developed by Hurtt and others using the same general methodology as Hurtt (2006) but use HYDE 3.1 historical data set for crop, pasture, and urban area (1500-2005).
Soil	UNASD (MsTMIP)	Static (One Time)	Static (One Time)	This is the MsTMIP driver for North America (Unified North American Soil Database) aggregated up to 0.5-degree resolution.
Biome	Derived from SYNMAP (MsTMIP)	Static (One Time)	Static (One Time)	SYNMAP biomes were simplified into 5 biomes for the PalEON domain and fractional cover of 10 common model plant functional types
Nitrogen deposition	MsTMIP (Enhanced Dentener extrapolation)	850-2010	Annual	

Model Spin-up & Initial Conditions

All simulations should be spun up to steady-state initial conditions prior to the start of the PalEON simulations through time. In all cases models should **use the oldest 20 years of the meteorological data for the spin-up period** (e.g., 850-869). CO2 concentrations should be set at 277 ppm, the mean of the 850-869 time period from the reconstructed CO2 concentration driver dataset. The use of a semi-analytical spinup (SAS) approach[1] is both permissible and encouraged. Please note that the additional variables required for the SAS and traceability framework[2] have been added to the protocol as encouraged outputs.

1[!] Xia, J. Y., Y.Q. Luo, Y.-P. Wang, E. S. Weng, O. Harauk. 2012. A semi-analytical solution to accelerate spin-up of a coupled carbon and nitrogen land model to steady state. Geoscientific Model Development 5:1259-1271.

2[!] Xia, J. Y., Y.Q. Luo, Y.-P. Wang, O. Harauk. 2013. Traceable components of terrestrial carbon storage capacity in biogeochemical models. Global Change Biology 19:2104-2116.

At the start of spin up, any prognostic soil temperatures should be initialized to the mean temperature of the 20 years of data that are being repeated. Canopy temperatures and canopy air space temperatures, pressures, and humidity should be initialized to the first observation in the met files. Biogeochemical pools and all other prognostic variables can be initialized as best suited for your model, however for models that possess a dynamics vegetation component, we strongly discourage the use of the SYNMAP-derived biome products, or other potential vegetation maps, as a number of our hypotheses will involve comparison to these data sets and we want the dynamic models to be independent when possible. Otherwise, models with static vegetation will use the SYNMAP-derived products to specify biome.

Model Outputs

Uploading Instructions:

All variables should be output as **monthly means** whenever possible. If your model is unable to provide monthly data, please provide annual resolution and indicate this as a deviation from the protocol in the README described below. **We also ask that you please fill update the model output spreadsheet with what the appropriate temporal resolution for each output variable from your model.**

Unless otherwise indicated, all variables should be packaged together into **100-year netCDF files** with dimensions of [lon, lat, time] or [lon, lat, PFT, time], which means that these variables will have dimensions of [30, 80, 120]. The first file should be 0850-0899, with 100-year increments after that. Due to the size and number of output files, if a variable cannot be output by your model feel free to drop it from the file, but please describe that in the README described below.

Model output should be uploaded to the "models" folder within "**phase2_model_output**" on iPlant. If your model does not currently have a folder, please create one and within that location create a new folder titled [model name].v1. If you have previously provided model output, please sequentially number your provided output according to the scheme listed above and *include a description of changes made from previous version in an updated README*.

Please zip or tar each set of simulations into an archive with the following format: [simulation acronym]_[first year in window].zip/tar.bz2. Please use leading zeros in file names where appropriate to facilitate file organization (e.g. ED2_0850.nc, ED2_0900.nc, ED2_1000.nc). **Please send Christy a quick email notification (crollinson@gmail.com) after you've uploaded files**, so we can perform some quick checks on model output to make sure variables are provided (or documented as not possible from your model) and check for gross errors such as potential unit inconsistencies.

README file (REQUIRED!):

In addition to model outputs groups need to submit a README file that contains, at a minimum, the following information. Please note that these README files are crucial to document deviations from this protocol as well as to provide the comprehensive authorship list.

- **Model information**: Full model name and version; Model acronym; Preferred/standard citations; URL; Is the code used in this analysis publicly available? Where?
- **Modeling team:** Names of those who contributed significantly to producing the PALEON runs; Address; Email; Phone; Brief description of each person's contribution. **This section will be used to create the PalEON MIP authorship list for resulting publications.**
- **Spin-up Protocol**: Please briefly describe the method used to achieve the initial model steady-state. For example, some models have used a semi-analytical solution (Xia *et al.* 2012) while others have done a multi-centennial "brute force" spin to achieve stead-state.
- **Modifications:** Any modifications to this protocol that you made in order to produce your runs. This may include any additional input data sets your model required; List any of the standard input data sets or variables that you model did NOT require; Document any deviations from the PalEON protocol. **Please make particular note of any modifications to the meteorological data necessary to drive the model.** For example, ED2 internally uses linear interpolation of the 6-hourly drivers whereas other models use their own temporal downscaling approach and others still require aggregation up to daily or monthly time steps.
- **Traceability information**: any constant annual transfer rates within your model for NPP allocation to plant carbon pools or between ecosystem carbon pools (Examples: fraction NPP allocated to wood, fraction slow SOM to passive SOM).
- **Model Settings:** Provide any settings files that detail the choices for model-specific conditions or settings as well as any available files that specify model parameters such as PFT characteristics or prescribed disturbance. (For example, the ED2IN files for ED2). If no settings file is available, please note the following settings in the ED2IN:
  - What modes of disturbance are used?
  - Is land use turned on or off?
  - Is there a nitrogen limitation scheme used in the runs?
  - Which PFTs are enabled?
- **Change Log – Changes made from previously provided output:** This may include running the model with updated met driver, altered model settings, or changes in the post-processing.

Requested Output Variables:

Please update variable availability from your model and temporal resolution on the master spreadsheet: *https://docs.google.com/spreadsheets/d/1f0LXVnHLglppztsFbYzDdaC5TMSZYUGg4OM__MnmaDM/edit?usp=sharing*

Category	Variable name	Long name	Units	Description
Diversity	PFT	PFT name	-	Name of each plant functional type or species included in the model. Dimensions: [PFT]
	Fcomp	Fractional Composition	kgC/kgC	**AGB** fractional composition of each PFT within each grid cell. Dimensions: [lon, lat, PFT, time]
	BA	Basal Area	m2/ha	Basal area by PFT Dimensions: [lon, lat, PFT, time]
	Dens	Stem density	1/ha	Stem Density by PFT Dimensions: [lon, lat, PFT, time]
	Estab	Establishment		New individuals Dimensions: [lon, lat, PFT, time]
	Mort	Mortality		Individuals lost through eath Dimensions: [lon, lat, PFT, time]
Carbon Pools	AGB	Aboveground biomass	kgC/m^2	Total aboveground biomass Dimensions: [lon, lat, time]
	TotLivBiom	Total living biomass		Total carbon content of living biomass (e.g. leaf +root+wood) Dimensions: [lon, lat, time]
	TotSoilCarb	Total soil carbon		Total soil and litter cabon content over the entire soil profile Dimensions: [lon, lat, time]
	CarbPools	Size of each carbon pool (e.g. leaf, wood, root, litter, CWD, soil)		However vegetation and soils are broken down in your model. Dimensions: [lon, lat, time, pool]
	poolname	Names of carbon pools		Names of each veg and soil carbon pool. Dimensions: [pool]
Carbon Fluxes	GPP	Gross primary productivity	$kgC/m^2/s$	Dimensions: [lon, lat, time]
	AutoResp	Autotrophic Respiration		Dimensions: [lon, lat, time]
	HeteroResp	Heterotrophic Respiration		Dimensions: [lon, lat, time]
	NPP	Net Primary Productivity		NPP of each PFT within each grid cell. **Dimensions: [lon, lat, PFT, time]**
	NEE	Net Ecosystem Exchange		Dimensions: [lon, lat, time]
	Fire	Fire Emissions		Dimensions: [lon, lat, time]
	GWBI	Gross woody biomass increment	kgC/m2/month	Variable most analogous to tree-ring-derived change in stem biomass (before mortality/CWD flux) Dimensions: [lon, lat, time]

	CWDI	Coarse Woody Debris Increment		Variable most analogous to flux of woody material to the detrital pool resulting from mortality; corresponds to GWBI Dimensions: [lon, lat, time]
Energy fluxes	LW_albedo	Longwave Albedo	-	Dimensions: [lon, lat, time]
	SW_albedo	Shortwave Albedo	-	Dimensions: [lon, lat, time]
	LWnet	Net Longwave Radiation	W/ m^2	Dimensions: [lon, lat, time]
	SWnet	Net Shortwave Radiation		Dimensions: [lon, lat, time]
	Qh	Sensible Heat		Dimensions: [lon, lat, time]
	Qle	Latent Heat		Dimensions: [lon, lat, time]
Other	LAI	Leaf Area Index	m^2/m^{-2}	Dimensions: [lon, lat, time]
	Qs	Surface runoff	$kg/m^2/s$	Dimensions: [lon, lat, time]
	Qsb	Subsurface runoff	$kg/m^2/s$	Drainage and subsurface lateral flow Dimensions: [lon, lat, time]
	Evap	Total Evaporation		Sum of evaporative sources minus transpiration Dimensions: [lon, lat, time]
	Transp	Total Transpiration		Transpiration of each PFT within each grid cell. **Dimensions: [lon, lat, PFT, time**]
	SnowDepth	Total snow depth		Dimensions: [lon, lat, time]
	SWE	Snow water equivalent	kg/m^2	Total water mass (ice plus liquid) Dimensions: [lon, lat, time]
	SoilMoist	Soil moisture	kg/m^2	Soil water content in each model-defined soil layer. **Dimensions: [lon, lat, nsoil, time]**
	SoilTemp	Soil temperature	K	Soil temperature in each model-defined soil layer. **Dimensions: [lon, lat, nsoil, time]**
	SoilDepth	Soil layer depths	m	Depth to the bottom of each model-defined soil layer: Dimensions: [nsoil]
Met drivers	lwdown	Incoming long-wave radiation	W m-2	Incident longwave radiation averaged over the time step of the forcing data Dimensions: [lon, lat, time]
	swdown	Incoming short-wave radiation	W m-2	Incident radiation in the shortwave part of the spectrum averaged over the time step of the forcing data Dimensions: [lon, lat, time]
	precipf	Precipitation	kg m-2 s-1	The per unit area and time precipitation representing the sum of convective rainfall, stratiform rainfall, and snowfall Dimensions: [lon, lat, time]
	psurf	Surface pressure	Pa	Pressure at the surface Dimensions: [lon, lat, time]

	qair	Specific humidity	kg kg-1	Specific humidity measured at the lowest level of the atmosphere Dimensions: [lon, lat, time]
	tair	Air temperature	K	2 meter near surface air temperature Dimensions: [lon, lat, time]
	wind	Wind speed	m s-1	Wind speed measured with a vertical coordinate in height of 10 m Dimensions: [lon, lat, time]
	CO2	CO2 concentration	ppm	Carbon dioxide concentration in the air Dimensions: [lon, lat, time]

Additional Information

Primary PalEON MIP Contacts (As of Fall 2015)

Name	**Email**	**Role**
Christy Rollinson	crollinson@gmail.com	PalEON modeling Postdoc (Boston U.)
Yao Liu	yaoliu@email.arizona.edu	PalEON modeling Postdoc (U. AZ)
Michael Dietze	dietze@bu.edu	PalEON modeling Co-PI (Boston U.)
Dave Moore	davidjpmoore@gmail.com	PalEON modeling Co-PI (U. AZ)
Jason McLachlan	jmclachl@nd.edu	PalEON lead PI (Notre Dame)
Jody Peters	peters.63@nd.edu	PalEON Program Manager (Notre Dame)