Data Inputs and Outputs of Demeter-W (Version 1.0)

Li, Xinya Xinya.Li@pnnl.gov

Hejazi, Mohamad I Mohamad.Hejazi@pnnl.gov

Huang, Zhongwei Zhongwei.Huang@pnnl.gov

Liu, Lu Lu.Liu@pnnl.gov

Feng, Leyang Leyang.Feng@pnnl.gov

September 09, 2017

Demeter-W can be download from https://github.com/JGCRI/demeter-w

The input data files needed by Demeter-W are categorized by different **folders**, which are introduced as follows:

1. ***demeter_w/reference***
   The files included in this folder are dependent files used by temporal downscaling modules of Demeter-W. They will be copied at install time to demeter-w's folder inside site-packages.

- *dist.csv*
  Function: This file lists the distances between a target basin and its neighbor basins, to show the closest neighbor basins for a target basin. The basin IDs are based on the global basin map (235 basins). It is a fixed input.
  File Format: csv, one header
  Dimension: 4 columns (target basin #, neighbor basin #, distance in m, value ID)
  Unit: meter
  Reference: N/A

- *obv_dom.csv*
  Function: Observed domestic water withdrawals (averaged) in 12 months at five cities in five different countries. The values do not need to be unformed under the same unit for different cities since normalized values (=each month / 12 month total) will be used by Demeter-W.
  File Format: csv, one-row header
  Dimension: 5 rows (5 cities) * 18 columns (city name, country name, period of years, gird index in 67420 cells, latitude, longitude, data of Jan, Feb, …, Dec)
  Unit: N/A
  Reference: Collected and consolidated by Huang et al. (2017)

- *IEA_9_Countries_Monthly_AvgElectricity_2000_2015.csv*

Function: Observed electricity generation (averaged) in 12 months at 9 countries. Normalized values (=each month / 12 month total) will be used by Demeter-W.
File Format: csv, one-row header
Dimension: 13 rows (12 months + 1) * 10 columns (9 countries + 1), the first row of data lists the country IDs from 249 counties (ID: 0-248), the first column of data lists the month index.
Unit: GWh
Reference: Collected and consolidated by Huang et al. (2017) from IEA historical data

2. *Example/Input*

The files included in this folder are defined in the configuration file and then imported into Demeter-W.

- *BasinNames.csv*
  Function: 235 basin names corresponding to *harmonized_inputs/basin.csv* in Table 1
  File Format: csv, no header
  Dimension: 235 row * 1 column
  Unit: N/A
  Reference: N/A

- *Grid_Areas_ID.csv*
  Function: Area value of each land grid cell
  File Format: csv, no header
  Dimension: 67420 rows * 1 column, decimal values
  Unit: ha (convert to km^2 by *0.01)
  Reference: N/A

- *coordinates.csv*
  Function: Coordinates of the 67420 grid cells
  File Format: csv, no header
  Dimension: 67420 rows * 5 columns (ID #, longitude, latitude, ilon, ilat)
  Unit: N/A
  Reference: N/A

- *country-names.csv*
  Function: 249 country names corresponding to *harmonized_inputs/country.csv* in Table 1
  File Format: csv, no header
  Dimension: 235 row * 2 column, first column is ID (integer, 0-248), second column is country name (string)
  Unit: N/A
  Reference: N/A

- ***GCAM***

- ***CSV/Case001***
  This folder contains the data files extracted from GCAM outputs (database in BaseX format). The csv files listed are summarized from the raw files in the folder "**Raw_From_GCAM**". "*Case001.txt*" described the relationships between the data files and the raw files. The example case has 32 regions and 5 years (2005, 2010, 2015, 2020, 2025).

- *query.xml* and *query_new.xml*
  These two files, providing rules, were used by gcam_reader (https://github.com/JGCRI/gcam_reader) to query data from GCAM database. The outputs are the files in "**Raw_From_GCAM**".

- ***TemporalDownscaling***
  - *DomesticR.csv*
    Function: The amplitude (dimensionless) which measures the relative difference of domestic water withdraw between the warmest and coldest months. It is used in the temporal downscaling of domestic sector.
    File Format: csv, one-row header
    Dimension: 67420 rows * 1, decimal values
    Unit: N/A
    Reference: calibrated by Huang et al., 2017

  - ***Irrigation***
    The monthly gridded irrigation water withdrawal was estimated by relying on monthly irrigation results from other global hydrological models to quantify monthly weighting profiles of how irrigation is spread out within a year in a particular region and per crop type. In this folder, three global gridded monthly irrigation water withdrawal models are provided as references for the period of 1971-2010:

    H08 (Hanasaki et al., 2008a; 2008b)
    h08_wfdei_hist_varsoc_pirrww_global_monthly_1971_2010.nc

    LPJmL (Rost et al., 2008)
    lpjml_wfdei_hist_varsoc_co2_pirrww_global_monthly_1971_2010.nc

    PCR-GLOBWB (Van Beek et al., 2011; Wada et al., 2011)
    pcrglobwb_wfdei_varsoc_pirrww_global_monthly_1971_2010.nc

    The original data files were obtained from ISI-MIP (Warszawski et al., 2014). We processed the original data files into gridded monthly percentage values as the weighting profiles applied in temporal downscaling of irrigation.

The data files are classic NetCDF file (http://www.unidata.ucar.edu/software/netcdf/docs/netcdf_introduction.html).
They follow the same format:
Size: 67420 x 480
Dimensions: index, month
Datatype: single
Variables: 'pirrww'

For more information about these three models and their data files, please refer to Huang et al., 2017.

- ***ClimateForcing/WATCH***
  For temporal downscaling of electricity and domestic water withdrawal from annual to monthly, the gridded daily air temperature data from WATCH forcing data methodology applied to Era Interim reanalysis data (WFDEI) from 1971 to 2010 is applied (Weedon et al., 2014).
  tas_watch_wfdei_monthly_1971_2010.mat
  Function: Averaged temperature in a month.
  File Format: mat (MATLAB formatted data), one variable: "tas"
  Dimension: 67420 * 480 (480 months in 1971 - 2010)
  Unit: Celsius

  hdd_cdd_tas_watch_wfdei_monthly_1971_2010.mat
  Function: HDD (heating degree days) and CDD (cooling degree days) in a month
  File Format: mat (MATLAB formatted data), two variables: "hdd" and "cdd"
  Dimension: 67420 * 480 (480 months in 1971 - 2010)
  Unit: Celsius

  For more information about how to calculate HDD and CDD, please refer to Huang et al., 2017.

- ***harmonized_inputs***
  The term "grid" is used to describe the spatial resolution of 0.5 geographic degrees. A global full data map contains a total of 259,200 grid cells (360 x 720) of which 67,420 grid cells are categorized as "land grids" and are considered valid for simulation purposes. In this study, the land grid cells are used to define a "gridded" map according to the coordinates and the indexes of the 67,420 cells on the 360x720 grid. The inputs converted using the 67,420 grid cells according to the coordinate data file are called harmonized inputs.
  Uniformed dimension and format in this folder: one-row header csv file, 67,420*1, **0 means no assignment**.
  The files are listed in Table 1.

- ***rgn32***

Global data map can be divided into 32 regions. The data files related to these 32 regions are saved in this folder. If other type of division (e.g. 14 regions, 235 basins) is used, prepare a similar folder contains all the listed data files.

- ***TD_Elec_paras***

  Five tables are included in this folder:

  ElecBuilding_1971_2010 P_b
  ElecBuildingHeat_1971_2010 P_h
  ElecBuildingCool_1971_2010 P_c
  ElecBuildingOthers_1971_2010 P_u
  ElecIndustry_1971_2010 P_{it}

  P_b and P_{it} are the proportions of total electricity use for building and transportation and industry together, respectively, $P_b + P_{it} = 1$.

  P_h, P_c and P_u are the proportions of total building electricity use for heating, cooling and other home utilities, respectively, $P_h + P_c + P_u = 1$.

  Function: The proportions of electricity use, are used in temporal downscaling of electricity sector
  File Format: csv, no header
  Dimension: 32 rows * 40 columns (32 regions * 40 years), decimal values (0-1)
  Unit: N/A
  Reference: IEA historical data, Huang et al. (2017)

- *RgnNames.csv*

  Function: The region names (e.g. "USA") and its corresponding id (e.g. 1 for USA).
  File Format: csv, one-row header
  Dimension: 32 rows * 2 columns, first column for region name (string), second column for region id (integer, 1-32)
  Unit: N/A
  Reference: N/A

- *bfracFAO2005.csv* and *gfracFAO2005.csv*

  Function: Livestock in GCAM has five sectors (beef, dairy, pork, poultry and sheepgoat), to be reorganized into six sectors used by Demeter-W (buffalo, cattle, goat, sheep, pig and poultry), two fraction numbers are needed.. One is recorded in bfracFAO2005.csv, the other is recorded in gfracFAO2005.csv:
  buffalo = (beef + diary)*bfrac
  cattle = (beef + diary)*(1-bfrac)
  goat = sheepgoat *gfrac
  sheep = sheepgoat *(1-gfrac)

  File Format: csv, one-row header
  Dimension: 32 rows * 2 columns, first column for region name (string), second column for buffalo (bfrac) or goat(gfrac) fraction (decimal, < 1) in this region

Unit: N/A
Reference: They are estimated from FAO gridded livestock of the world (refer to "*harmonized inputs/livestock_ *.csv*")

- *irrigation-frac.csv*
  Function: Pre-calculated irrigation shares as a fixed input, is used when extracting data files from GCAM database, if GCAM didn't produce endogenous irrigated and rain-fed land allocations.
  File Format: csv, one-row header
  Dimension: 26 columns: region #, aez #, crop #, 1990, 2005, 2010, ..., 2100, region name, crop name, decimal values (0-1) for column 4-24
  Unit: N/A
  Reference: N/A

The output data files generated by Demeter-W are introduced as follows:

Example/ Output/Test001

The files included in this folder are example output files.

logfile.log

This is the log file of a simulation, recording project settings, simulation progress, warnings and errors, summarized diagnostics information and CPU times, etc. The following example message will show at the end of a log file when simulation runs successfully:

```
Saving outputs...
Save the gridded water usage results for each withdrawal category in
NetCDF format (Unit: km3/yr)
Save the monthly water usage results for each withdrawal category
(Unit: km3/month)
---Output: 1.66400003433 seconds ---
('End Project:   ', 'Test001')
```

Diagnostics_Spatial_Downscaling.csv

This table is created by diagnostics module for spatial downscaling. It lists the water withdrawal values for each year, each region and each water demand sector comparing between downscaled results and input data from GCAM along with their differences.

Diagnostics_Temporal_Downscaling_Irrigation.csv

Diagnostics_Temporal_Downscaling_Domestic.png

Diagnostics_Temporal_Downscaling_Electricity.png

These three files are created by diagnostics module for temporal downscaling. A diagnostics table for irrigation sector will be generated listing the water withdrawal values for each year, each basin before and after temporal downscaling. Two figures will also be plotted to show the simulated monthly profiles for domestic sector and electricity generation sector respectively, comparing with overserved results. The observed monthly profiles were imported from files described in "***demeter_w/reference***" (Huang et al., 2017).

The downscaled withdrawal results for each sector will be saved. They are divided into two groups, Spatial Downscaling (SD) results (the file name starts with "wd") and Temporal Downscaling (TD) results (the file name starts with "twd"):

Sector	SD Results	TD results
Domestic	wddom	twddom
Electricity Generation	wdelec	twdelec
Irrigation	wdirr	twdirr
Livestock	wdliv	twdliv
Manufacturing	wdmfg	twdmfg
Mining	wdmin	twdmin
Non-Agriculture	wdnonag	-
Total	wdtotal	-

There are two formats: nc (classic netcdf) and csv.

For example, in example case, the downscaled results were saved in nc format.

For SD results, they follow the same structure:

Size: 67420, 5
Dimensions: index, year
Datatype: single (float32)
Variables (dimension):
data (index, year)
ID (index)
lon (index)
lat (index)
ilon (index)
ilat (index)
mapindex (index)
years (year)

For TD results, they follow the same structure:

Size: 67420, 72
Dimensions: index, month
Datatype: single (float32)
Variables (dimension):
- data (index, month)
- ID (index)
- lon (index)
- lat (index)
- ilon (index)
- ilat (index)
- mapindex (index)
- months (month)

csv files will save the same variables into tables.

Table 1

File Name	Description	Reference	Unit
AEZ	AEZ ID for each cell, 18 zones: 1-18,	-	-
basin	Basin ID for each cell, 235 basins: 1-235	-	-
country	Country ID for each cell, 249 countries: 1-249	-	-
GMIA_cropland	Irrigation areas in 2005 in each grid cell	Siebert, 2013	km^2
HYDE_cropland	Irrigation area in 1900-2000 in each grid cell: every 10 years	HYDE 3.1 Final, Klein Goldewijk et al., 2011	km^2
HYDE_grassland	Irrigation areas in 1900-2000 in each grid cell: every 10 years	HYDE 3.1 Final, Klein Goldewijk et al., 2011	km^2
GPW_population	Population: 1990-2015 data, every 5 years	CIESIN, 2016	-
HYDE_population	Population: 1750-2000 data, every 10 years	HYDE 3.1 Final, Klein Goldewijk et al., 2011	-
livestock_buffalo	Number of buffalo in each grid cell	Wint and Robinson, 2007.	head
livestock_cattle	Number of cattle in each grid cell	Wint and Robinson, 2007.	head
livestock_goat	Number of goat in each grid cell	Wint and Robinson, 2007.	head
livestock_pig	Number of pig in each grid cell	Wint and Robinson, 2007.	head
livestock_poultry	Number of poultry in each grid cell	Wint and Robinson, 2007.	head
livestock_sheep	Number of sheep in each grid cell	Wint and Robinson, 2007.	head
soil_moisture	Maximum Soil Moisture	FAO, 2003.	mm/month
region32_grids	Region ID for each cell, 32 regions: 1-32	-	-

References

Huang, Z., Hejazi, M., Li, X., Tang, Q., Leng, G., Liu, Y., Döll, P., Eisner, S., Gerten, D., Hanasaki, N., and Wada, Y., 2017. Reconstruction of global gridded monthly sectoral water withdrawals for 1971–2010 and analysis of their spatiotemporal patterns, *Hydrology and Earth System Sciences Discussions*, DOI: https://doi.org/10.5194/hess-2017-551

Wada, Y., Van Beek, L.P.H., Viviroli, D., Dürr, H.H., Weingartner, R. and Bierkens, M.F., 2011. Global monthly water stress: 2. Water withdrawal and severity of water stress. Water Resources Research 47(7): W07518. DOI: http://dx.doi.org/10.1029/2010WR009792

Hanasaki, N., Kanae, S., Oki, T., Masuda, K., Motoya, K., Shirakawa, N., Shen, Y. and Tanaka, K., 2008a. An integrated model for the assessment of global water resources–Part 1: Model description and input meteorological forcing. Hydrology and Earth System Sciences 12(4): 1007-1025. DOI: https://doi.org/10.5194/hess-12-1007-2008

Hanasaki, N., Kanae, S., Oki, T., Masuda, K., Motoya, K., Shirakawa, N., Shen, Y. and Tanaka, K., 2008b. An integrated model for the assessment of global water resources–Part 2: Applications and assessments. Hydrology and Earth System Sciences 12(4): 1027-1037. DOI: https://doi.org/10.5194/hess-12-1027-2008

Rost, S., Gerten, D., Bondeau, A., Lucht, W., Rohwer, J. and Schaphoff, S., 2008. Agricultural green and blue water consumption and its influence on the global water system. Water Resources Research 44(9): W09405. DOI: https://doi.org/10.1029/2007WR006331

Van Beek, L.P.H., Wada, Y. and Bierkens, M.F., 2011. Global monthly water stress: 1. Water balance and water availability. Water Resources Research 47(7): W07517. DOI: https://doi.org/10.1029/2010WR009791

Klein Goldewijk, K., Beusen, A., Van Drecht, G. and De Vos, M., 2011. The HYDE 3.1 spatially explicit database of human induced global land use change over the past 12,000 years. Global Ecology and Biogeography 20(1): 73-86. DOI: https://doi.org/10.1111/j.1466-8238.2010.00587.x. http://beta.sedac.ciesin.columbia.edu/data/set/gpw-v4-population-count

Center for International Earth Science Information Network (CIESIN) - Columbia University. 2016. Gridded Population of the World, Version 4 (GPWv4): Population Count. NASA Socioeconomic Data and Applications Center (SEDAC), Palisades, NY. DOI: http://dx.doi.org/10.7927/H4X63JVC. http://beta.sedac.ciesin.columbia.edu/data/set/gpw-v4-population-count

Siebert, S., Henrich, V., Frenken, K., and Burke, J., 2013. Global Map of Irrigation Areas version 5. Rheinische Friedrich-Wilhelms-University, Bonn, Germany / Food and Agriculture Organization of the United Nations, Rome, Italy. http://www.fao.org/nr/water/aquastat/irrigationmap/index10.stm

Wint, W. and Robinson, T., 2007. Gridded livestock of the world. Food and Agriculture Organization (FAO), report 131, Rome. http://www.fao.org/docrep/010/a1259e/a1259e00.htm

Food and Agriculture Organization of the United Nations (FAO), 2003. Digital soil map of the world and derived soil properties [Version 3.5], FAO, Rome, Italy.

Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O. and Schewe, J., 2014. The inter-sectoral impact model intercomparison project (ISI–MIP): project framework. *Proceedings of the National Academy of Sciences*, *111*(9), pp.3228-3232.