User's Guide for China Meteorological Forcing Dataset

(Data version: 01.06.0014)

Kun Yang, Prof., Dr. (yangk@itpcas.ac.cn)
Data Assimilation and Modeling Center for Tibetan Multi-spheres (DAM)
Institute of Tibetan Plateau Research (ITP), Chinese Academy of Sciences (CAS)

1. Introduction

The China Meteorological Forcing Dataset (CMFD) was produced by merging a variety of data sources. This dataset currently covers the period 1979-2018. Its spatial resolution is 0.1 deg. and its temporal resolution is 3-hr. The data set can be used for hydrological modeling, land surface modeling, land data assimilation, and other terrestrial modeling.

This data sources used to produce the forcing data include:

(1) CMA (China Meteorological Administration) weather station observation data (wind, air temperature, relative humidity, sunshine duration, precipitation, and surface pressure) for the period of 1979-2018. The observed data together with the radiation data estimated from the observed sunshine duration are used as true values of the meteorological parameters.

(2) TRMM satellite precipitation analysis data (3B42) for the period of 1998-2018 and GLDAS precipitation for the period of 1979-2018. The GLDAS precipitation data during 1998-2018 were only used to replace TRMM 3B42 data that are not available beyond 40°N.

(3) GEWEX-SRB downward shortwave radiation for the period of 1983.07-2007.12, and GLDAS downward shortwave radiation data for the period of 1979.01-1983.06 and 2008.01-2018.12.

(4) The Modern Era-Retrospective Analysis for Research and Applications (MERRA) (surface pressure) for the period of 1979-2015. GLDAS data (surface pressure) were used after 2015.

(5) GLDAS data (wind, air temperature, relative humidity) for the period of 1979-2018.

The data fusion technique is to be introduced in a manuscript in preparation.

2. Variables

Variables	Variable name	Unit	Physical meaning
Temperature	temp	K	Instantaneous near surface (2 m) air temperature.
Pressure	pres	Pa	Instantaneous near surface (2 m) air pressure.
Specific humidity	shum	kg kg^{-1}	Instantaneous near surface (2 m) air specific humidity.
Wind speed	wind	m s^{-1}	Instantaneous near surface (10 m) wind speed.
Downward shortwave radiation	srad	W m^{-2}	3-hourly mean (from -1.5 hr to +1.5 hr) surface downward shortwave radiation.
Downward longwave radiation	lrad	W m^{-2}	3-hourly mean (from -1.5 hr to +1.5 hr) surface downward longwave radiation.
Precipitation rate	prec	mm hr^{-1}	3-hourly mean (from -3.0 hr to 0.0 hr) precipitation rate.

3. Data archive and access

This dataset can be obtained at

http://data.tpdc.ac.cn/en/data/8028b944-daaa-4511-8769-965612652c49/ .

The data archive is arranged as follows:

(1) Each data file is named as *VVVV*_CMFD_V*vvvv*_*CCCC*_*YYYYMM*.nc (e.g. temp_CMFD_V0105_B-01_197901.nc), where *VVVV* is the abbreviation of variable name, *vvvv* is the data version, *CCCC* is the category of dataset, *YYYY* is the four-digital year, *MM* is the two-digital month, that means one file contains data of only one variable for only one month, additionally, .nc is the file name suffix which indicates that the data were stored as NetCDF format;

(2) Each file is compressed as .gz format in order to save disk space; users have to uncompress these data files before using them. Recommended software to do this is gunzip (on UNIX/Linux) or WinRAR (on Windows);

(3) Another data-compressing method is used in NetCDF files via packing 32-bits float values to 16-bits short integer. Two parameters, scale_factor and add_offset, are crucial to this technic. Most of high-level software (e.g. GrADS, Ferret) can automatically unpack data using these two parameters so users don't need to care them, but when users want to write programs themselves to read data from NetCDF files, the following formula should be used to unpack 16-bits short integer data to 32-bits float:

unpacked_data_value = packed_data_value * scale_factor + add_offset

The following table is scale_factor and add_offset of each variables.

Variables	Variable name	Scale_factor	Add_offset
Temperature	temp	0.01	273.15
Pressure	pres	2.00	63500.00
Specific humidity	shum	0.000001	0.025
Wind speed	wind	0.002	60.00
Downward shortwave radiation	srad	0.25	685.00
Downward longwave radiation	lrad	0.25	685.00
Precipitation rate	prec	0.0025	50.00

(4) A simple FORTRAN program named PRG-01.01_Data_Read_Example.f90 is provided along with the dataset as a sample to show how to read the NetCDF data files, but users might have to modify it to satisfy special needs. This program has been tested on Linux and Windows platform with netcdf-3.x libraries, and anyone who wants to compile this program must have NetCDF library installed first. For more information about NetCDF, please see http://www.unidata.ucar.edu/software/netcdf.

(5) Two mask files named mask_ITPCAS-CMFD_V0106_B-01_010deg.nc and mask_ITPCAS-CMFD_V0106_B-01_010deg_with_lake.nc are also provided along with this dataset, which has same grid distribution as the data product. In this file, grid points locate at China's land give values of 1, while the other grid points give values of 0. The only difference of the two mask files is that one of

them treat lake as land while the other one does not. Another file named elev_ITPCAS-CMFD_V0106_B-01_010deg.nc is terrain elevation.

4. Caution of usage (Known problems)

The downward shortwave radiation is developed using more than one dataset as background field. As stated in section 1, we used GLDAS data (downward shortwave radiation) for the period of 1979.01-1983.06 and 2008.01-2010.12, and GEWEX-SRB downward shortwave radiation during 1983.07-2007.12. This would have probably generated inconsistency in the data time serials. Additionally, this problem also exists in precipitation data product and downward longwave radiation (because downward longwave radiation is calculated using the downward shortwave radiation and air temperature and humidity products).

5. License

This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/ or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

6. Data Acknowledgement

(1) A formal citation of the data, whenever the ITP forcing data are being used for publication of scientific results, is as follows:

YANG Kun. China Meteorological Forcing Data (1979-2018). Big Data System for Pan-Third Pole, 2018. doi: 10.11888/AtmosphericPhysics.tpe.249369.file

Regarding the radiation estimation, you may refer to:

Kun Yang, Jie He, Wenjun Tang, Jun Qin, CCK Cheng, 2010: On downward shortwave and longwave radiations over high altitude regions: Observation and modeling in the Tibetan Plateau. Agric. Forest. Meteorol, 150, 38-46.

(2) To recognize the valuable role of the data provider, the data's origin must be acknowledged ("*The forcing dataset used in this study was developed by Data Assimilation and Modeling Center for Tibetan Multi-spheres, Institute of Tibetan Plateau Research, Chinese Academy of Sciences*").

(3) To facilitate the maintenance and improvements of the dataset, please send a copy of scientific publications to the data provider via email (**yangk@itpcas.ac.cn**).

7. Disclaimer

Although we have made great efforts to control input data and improve the forcing dataset, the accuracy and reliability of these data is subject to the availability of in situ data and is not guaranteed in any way. The provider disclaims liability of any kind whatsoever, including, without limitation, liability for quality, performance, and fitness for a particular purpose arising out of the use.