NASA Earth Exchange
Global Daily Downscaled Projections (NEX-GDDP)

1. Intent of This Document and POC

1a) This document provides a brief overview of the NASA Earth Exchange (NEX) Global Daily Downscaled Projections (GDDP) dataset, and is intended for users who wish to apply the NEX-GDDP dataset in studies of climate change impacts. This document summarizes essential information needed for accessing and using information contained within the NEX-GDDP dataset. References and additional information are provided at the end of this document

This NASA dataset is provided to assist the science community in conducting studies of climate change impacts at local to regional scales, and to enhance public understanding of possible future climate patterns at the spatial scale of individual towns, cities, and watersheds. This dataset is intended for use in scientific research only, and use of this dataset for other purposes, such as commercial applications and engineering or design studies is not recommended without consultation with a qualified expert. Community feedback to improve and validate the dataset for modeling usage is appreciated.

Email comments to bridget@climateanalyticsgroup.org with copy to forrest.s.melton@nasa.gov.

Dataset File Name: NASA Earth Exchange Global Daily Downscaled Projections (NEX-GDDP), https://nex.nasa.gov/nex/projects/1356/

1b) Technical points of contact for this dataset:

Dr. Bridget Thrasher, bridget@climateanalyticsgroup.org

Dr. Rama Nemani, rama.nemani@nasa.gov

2. Data Field Descriptions

CF variable name, units:	*tasmin* Daily Minimum Near-Surface Air Temperature Degrees Kelvin
Spatial resolution:	0.25 degrees x 0.25 degrees
Temporal resolution and extent:	Daily from 1950-01-01 00:00:00 to 2099-12-31 11:59:59 Units are in days since a reference date. The reference date varies by model, and is based on the reference date used in the corresponding CMIP5 GCM experiment.

CF variable name, units:	*tasmax* Daily Maximum Near-Surface Air Temperature Degrees Kelvin
Spatial resolution:	0.25 degrees x 0.25 degrees
Temporal resolution and extent:	Daily from 1950-01-01 00:00:00 to 2099-12-31 11:59:59 Units are in days since a reference date. The reference date varies by model, and is based on the reference date used in the corresponding CMIP5 GCM experiment.

CF variable name, units:	*pr* Precipitation (mean of the daily precipitation rate) kg m^{-2} s^{-1}
Spatial resolution:	0.25 degrees x 0.25 degrees
Temporal resolution and extent:	Daily from 1950-01-01 00:00:00 to 2099-12-31 11:59:59 Units are in days since a reference date. The reference date varies by model, and is based on the reference date used in the corresponding CMIP5 GCM experiment.

Dataset projection:	Geographic
Dataset datum:	WGS-84
Location of pixel *lat* and *long*	The pixel *lat* and *long* fields in the metadata provide the location of the center of each pixel
Coverage:	West Bounding Coordinate: 180° W East Bounding Coordinate: 180° E North Bounding Coordinate: 90° N South Bounding Coordinate: 90° S

3. Data Origin and Methods

3.1. Introduction

The NEX-GDDP dataset is comprised of downscaled climate scenarios for the globe that are derived from the General Circulation Model (GCM) runs conducted under the Coupled Model Intercomparison Project Phase 5 (CMIP5) [Taylor et al. 2012] and across two of the four greenhouse gas emissions scenarios known as Representative Concentration Pathways (RCPs) [Meinshausen et al. 2011]. The CMIP5 GCM runs were developed in support of the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC AR5). This dataset includes downscaled projections from the 21 models and scenarios for which daily scenarios were produced and distributed under CMIP5. The purpose of these datasets is to provide a set of global, high resolution, bias-corrected climate change projections that can be used to evaluate climate change impacts on processes that are sensitive to finer-scale climate gradients and the effects of local topography on climate conditions.

The demand for downscaling of GCM outputs arises from two primary limitations inherent with current global simulation results. First, most GCMs are run using relatively coarse resolution grids (e.g., a few degrees or 10^2km), which limit their ability to capture the spatial details in climate patterns that are often required or desired in regional or local analyses. Second, even the most advanced GCMs may produce projections that are globally accurate but locally biased in their statistical characteristics (i.e., mean, variance, etc.) when compared with observations.

The Bias-Correction Spatial Disaggregation (BCSD) method used in generating the NEX-GDDP dataset is a statistical downscaling algorithm specifically developed to address these current limitations of global GCM outputs [Wood et al. 2002; Wood et al. 2004; Maurer et al. 2008; Thrasher et al. 2012]. The algorithm compares the GCM outputs with corresponding climate observations over a common period and uses information derived from the comparison to adjust future climate projections so that they are (progressively) more consistent with the historical climate records and, presumably, more realistic for the spatial domain of interest. The algorithm also utilizes the spatial detail provided by observationally-derived datasets to interpolate the GCM outputs to higher-resolution grids.

With the help of the computational resources provided by NEX and the NASA Advanced Supercomputing (NAS) facility, we have applied the BCSD method to produce a dataset of downscaled CMIP5 climate projections to facilitate the assessment of climate change impacts in the United States. The dataset compiles 42 climate projections from 21 CMIP5 GCMs (Table 1) and two RCP scenarios (RCP 4.5 and RCP 8.5) for the period from 2006 to 2100, as well as the historical experiment for each model for the period from 1950-2005. Each of these climate projections is downscaled at a spatial resolution of 0.25 degrees x 0.25 degrees (approximately 25 km x 25 km) resulting in a data archive size of more than 11TB (1TB = 10^{12} Bytes).

This document provides a basic description of the implementation of the BCSD method as applied in the downscaling of the CMIP5 GCM data. Additional technical details for the algorithm may also be found in Wood et al. [2002, 2004], and Maurer et al. [2008]. The approach used to produce the NEX-GDDP dataset was previously applied to data from the CMIP3 archive, and the approach used in production of both datasets is described in detail in Thrasher et al. [2012].

3.2 Methods

3.2.1 Datasets

Climate Model Data: We compiled 42 climate projections from the 21 CMIP5 GCM simulations (Table 1) and two RCP scenarios (RCP 4.5 and RCP 8.5). Each of the climate projections includes daily average maximum temperature, minimum temperature, and precipitation for the periods from 1950 through 2005 ("Retrospective Run") and from 2006 to 2100 ("Prospective Run"). During the downscaling process, the retrospective simulations serve as the training data, and are compared against the observational climate records (see below). The relationships derived from the comparison are then applied to downscale the prospective climate projections.

Because all 42 climate projections are downscaled through the same procedures, for simplicity we refer to them as "GCM data" without differentiating any individual models.

Observational Climate Data: We use a climate dataset from the Global Meteorological Forcing Dataset (GMFD) for Land Surface Modeling, available from the Terrestrial Hydrology Research Group at Princeton University [Sheffield et al. 2006]. This dataset blends reanalysis data with observations and is currently available at spatial resolutions of 0.25 degrees, 0.5 degrees and 1.0 degree, and temporal resolutions of 3-hourly, daily, and monthly timesteps. For development of the NEX-GDDP dataset, we used the 0.25-degree, historical data for daily maximum temperature, daily minimum temperature, and daily precipitation from 1950 to 2005.

3.2.2 Data Pre-processing

Since the BCSD method does not explicitly adjust the trends (the slopes, in particular) in climate variables produced by GCMs, we extract the monthly large-scale climate trends from the GCM temperature data. This is calculated as a 9-year running average for each individual month (e.g. the trend for all Januaries taken together). These trends are preserved and added back to the adjusted data after the bias-correction step.

3.2.3 Bias Correction (BC)

The Bias-Correction step "corrects" the bias of the GCM data through comparisons performed against the GMFD historical data. For each climate variable in a given day, the algorithm generates the cumulative distribution function (CDF) for the GMFD data and for the retrospective GCM simulations, respectively, by pooling and sorting the corresponding source values (day of year +/- 15 days) over the period from 1950 through 2005. It then compares the two CDFs at various probability thresholds to establish a quantile map between the GCM data and the historical climate data. Based on this map, GCM values in any CDF quantile (e.g., p=90%) can be translated to corresponding GMFD values in the same CDF quantile. Assuming that the CDF of the GCM simulations is stable across the retrospective and the prospective periods, to "correct" the projected future climate variations the algorithm simply looks up the probability quantile associated with the predicted climate values from the estimated GCM CDF, identifies the corresponding observed climate values at the same probability quantile in the GMFD CDF, and then accepts the latter as the adjusted climate predictions. The climate projections adjusted in this way have the same CDF as the GMFD data; therefore, the possible biases in the statistical structure (the variance, in particular) of the original GCM outputs are removed by this procedure. At the end of the Bias-

Correction step, the previously extracted temperature climate trends are added back to the adjusted GCM climate fields

3.2.4 Spatial Disaggregation (SD)

The Spatial-Disaggregation step spatially interpolates the Adjusted GCM data to the finer resolution grid of the 0.25-degree GMFD data. Other than simple linear spatial interpolation, multiple steps are adopted in the SD algorithm to preserve spatial details of the observational data. First, the multi-decade daily climatology of the GMFD variables (temperature and precipitation) are generated at both native and GCM resolutions. The climatology for the SD step is the average for each day of the year calculated over the reference period, 1950-2005. Second, for each time step, the algorithm compares the Adjusted GCM variables with the corresponding GMFD climatology to calculate "scaling factors". In particular, the scaling factors are calculated as the differences between the bias-corrected GCM and the GMFD data for temperature, but as the quotients (between the two datasets) for precipitation to avoid negative values for the latter. Third, the coarse-resolution scaling factors are bilinearly interpolated to the fine-resolution GMFD grid. Finally, the scaling factors are applied, by addition or "shifting" for temperatures and by multiplication for precipitation, on the fine-resolution GMFD climatology to obtain the desired downscaled climate fields. As such, the algorithm essentially merges the observed historical spatial climatology with the relative changes at each time step simulated by the GCMs to produce the final results.

4. Considerations and Recommended Use

4.1 Recommended Use

This dataset has been generated and is being distributed to assist the science community in conducting studies of climate change impacts at local to regional scales, and to enhance public understanding of possible future climate patterns and climate impacts at the scale of individual cities, communities, and watersheds. This dataset is intended for use in scientific research only, and use of this dataset for other purposes, such as commercial applications, and engineering or design studies is not recommended without consultation with a qualified expert.

4.2 Assumptions and Limitations

The BCSD approach used in generating this downscaled dataset inherently assumes that the relative spatial patterns in temperature and precipitation observed from 1950 through 2005 will remain constant under future climate change. Other than the higher spatial resolution and bias correction, this dataset does not add information beyond what is contained in the original CMIP5 scenarios, and preserves the frequency of periods of anomalously high and low temperature or precipitation (i.e., extreme events) within each individual CMIP5 scenario.

4.3 Trend Adjustment to Individual Models

As described in Section 2.1, the BCSD algorithm does not adjust the *slope* of the trends in the GCM projections. In the case of temperature, for instance, if the GCM predicts a mean temperature increase of 2°C between 2006 and 2100, the same temperature change (i.e., a trend of 2°C over 95 years) will be observed in the downscaled temperature field. However, the BCSD algorithm does adjust the *offset* of the climate trends by shifting the retrospectively simulated climate variables (1950 through 2005) to match the GMFD data. In the previous example, if the simulated mean temperature from the GCM over the period 1996-2005 is 14°C, while the observed mean temperature is 15°C, the BCSD algorithm will correct the "bias" by shifting the GCM retrospective and prospective projections upward by 1°C. The adjusted mean temperature projected for the end of the 21st century will then be raised from 16°C to 17°C, though its relative change over the period 2006-2100 is preserved as 2°C. Though such adjustments of future climate projections are qualitatively justifiable, quantitatively the linear shifting itself may not be realistic because the climate system is nonlinear in nature. Users of this dataset should be aware of this limitation of the downscaled data, particularly when using downscaled scenarios from individual GCMs.

5. Credits and Acknowledgements

Please use the reference below as the primary citation for the methods used to produce this dataset:

Thrasher, B., Maurer, E. P., McKellar, C., & Duffy, P. B., 2012: Technical Note: Bias correcting climate model simulated daily temperature extremes with quantile mapping. *Hydrology and Earth System Sciences*, 16(9), 3309-3314.

Please add the following acknowledgement to any publications that result from use of this dataset:

> Climate scenarios used were from the NEX-GDDP dataset, prepared by the Climate Analytics Group and NASA Ames Research Center using the NASA Earth Exchange, and distributed by the NASA Center for Climate Simulation (NCCS).

6. References

Daly, C., R.P. Neilson, and D.L. Phillips, 1994: A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology*, **33**, 140-158.

Maurer, E. P. and Hidalgo, H. G., 2008: Utility of daily vs. monthly large-scale climate data: an intercomparison of two statistical downscaling methods, *Hydrology and Earth System Sciences*, **12**, 551-563.

Meinshausen, M. S.J. Smith, K. Calvin, J.S. Daniel, M.L.T. Kainuma, and et al., 2011: The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change*, **109**, 213-241.

Sheffield, J., G. Goteti, and E. F. Wood, 2006: Development of a 50-yr high-resolution global dataset of meteorological forcings for land surface modeling, J. Climate, 19 (13), 3088-3111.

Shepard, D.S., 1984: Computer mapping: The SYMAP interpolation algorithm, in Spatial Statistics and Models, edited by G.L. Gaile and C.J. Willmott. D. Reidel Publishing Co., Norwell, Dordrecht, Holland, pp. 133-145.

Taylor, Karl E., Ronald J. Stouffer, Gerald A. Meehl, 2012: An Overview of CMIP5 and the Experiment Design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498.

Thrasher, B., Xiong, J., Wang, W., Melton, F., Michaelis, A., & Nemani, R., 2013: Downscaled climate projections suitable for resource management. Eos, Transactions American Geophysical Union, 94(37), 321-323.

Thrasher, B., Maurer, E. P., McKellar, C., & Duffy, P. B., 2012: Technical Note: Bias correcting climate model simulated daily temperature extremes with quantile mapping. Hydrology and Earth System Sciences, 16(9), 3309-3314.

Wood, A.W., E.P. Maurer, A. Kumar, and D.P. Lettenmaier, 2002: Long-range experimental hydrologic forecasting for the eastern United States. *J. Geophysical Research-Atmospheres*, **107**, 4429, doi:10.1029/2001JD000659.

Wood, A.W., L.R. Leung, V. Sridhar, and D.P. Lettenmaier, 2004: Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic Change*, **15**,189-216.

7. Dataset and Document Revision History

Rev 1 – 23 July 2021 – Document created. This is a new document/dataset.

Table 1. CMIP5 models included in downscaled archive

ACCESS1-0	CSIRO-MK3-6-0	MIROC-ESM
BCC-CSM1-1	GFDL-CM3	MIROC-ESM-CHEM
BNU-ESM	GFDL-ESM2G	MIROC5
CanESM2	GFDL-ESM2M	MPI-ESM-LR
CCSM4	INMCM4	MPI-ESM-MR
CESM1-BGC	IPSL-CM5A-LR	MRI-CGCM3
CNRM-CM5	IPSL-CM5A-MR	NorESM1-M

Appendix I – Working with netCDF files

To work with the NEX-GDDP netCDF files you will need to have the netCDF libraries installed (https://www.unidata.ucar.edu/software/netcdf/docs/netcdf-install.html). If you are installing and building the libraries, be sure to include the ncdump utility.

Once the libraries are installed, you can use **ncdump** to get metadata information using the **–h** option. (Tip: It's very important not to forget to include **–h** when using the **ncdump** command.)

this command will display the metadata contained in the netCDF header for each file
% ncdump –h *filename*

For users who prefer a GUI interface, **ncbrowse** is a useful tool for browsing both metadata and data contents of netCDF files (http://www.epic.noaa.gov/java/ncBrowse/).

The Python netCDF4 libraries (http://code.google.com/p/netcdf4-python/) contain a number of highly useful functions for working with netCDF files. **numpy** (http://www.numpy.org/) is also highly recommended and provides a number of very useful statistical functions.

Once these libraries are installed, the following commands will be useful for working with the CMIP5 netCDF files in python.

Python commands:

to import the modules
% import netCDF4,numpy

to open a netCDF data file
the second argument 'r' means readonly, use 'a' to append/modify a file

% ds = netCDF4.Dataset(infilename,'r')

the 'variables' function will list the variables that are in the file
% ds.variables

to retrieve information on the shape of a specific variable (in this case, 'pr' or precipitation)
% ds.variables['pr'].shape

to retrieve first timestep from that variable
% pr0 = ds.variables['pr'][0]

to determine what value was used as a fill value for the variable
% pr.fill_value

to determine the minimum value

```
% numpy.min(pr0)
# to determine the maximum value
% numpy.max(pr0)

# to extract a subset of the full dataset contained in the file
% prSub = pr0[10:50,20:30]

# to close the file
% ds.close()
```