NEX-DCP30: Downscaled 30 Arc-Second CMIP5 Climate Projections for Studies of Climate Change Impacts in the United States

1. Intent of This Document and POC

1a) This document provides a brief overview of the NASA Earth Exchange (NEX) Downscaled Climate Projections (NEX-DCP30) dataset for the conterminous U.S., and is intended for users of the dataset who wish to apply the NEX-DCP30 dataset in studies of climate change impacts. This document summarizes essential information needed for accessing and using information contained within the NEX-DCP30 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 and climate impacts at the scale of individual neighborhoods and communities. 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.

Dataset File Name: NASA Earth Exchange (NEX) Downscaled Climate Projections (NEX-DCP30), https://portal.nccs.nasa.gov/portal_home/published/NEX.html

1b) Technical points of contact for this dataset:

Dr. Bridget Thrasher, bridget@climateanalyticsgroup.org
Forrest Melton, forrest.s.melton@nasa.gov

2. Data Field Descriptions

CF variable name, units:	*tasmax* Monthly Average Maximum Near-Surface Air Temperature (monthly mean of the daily minimum near-surface air temperature) Degrees Kelvin
Spatial resolution:	30 arc-seconds x 30 arc-seconds (0.0083333333 degrees x 0.0083333333 degrees)
Temporal resolution and extent:	Monthly from 1950-01 to 2099-12 Units are in days since 1950-01-01 00:00:00
Coverage:	West Bounding Coordinate: -125.02083333 East Bounding Coordinate: -66.47916667 North Bounding Coordinate: 49.9375 South Bounding Coordinate: 24.0625

CF variable name, units:	*tasmin* Monthly Average Minimum Near-Surface Air Temperature (monthly mean of the daily minimum near surface air temperature) Degrees Kelvin
Spatial resolution:	30 arc-seconds x 30 arc-seconds (0.0083333333 degrees x 0.0083333333 degrees)
Temporal resolution and extent:	Monthly from 1950-01 to 2099-12 Units are in days since 1950-01-01 00:00:00
Coverage:	West Bounding Coordinate: -125.02083333 East Bounding Coordinate: -66.47916667 North Bounding Coordinate: 49.9375 South Bounding Coordinate: 24.0625

CF variable name, units:	*pr* Monthly Average Precipitation (monthly mean of the daily precipitation rate) kg m-2 s-1
Spatial resolution:	30 arc-seconds x 30 arc-seconds (0.0083333333 degrees x 0.0083333333 degrees)
Temporal resolution and extent:	Monthly from 1950-01 to 2099-12 Units are in days since 1950-01-01 00:00:00
Coverage:	West Bounding Coordinate: -125.02083333 East Bounding Coordinate: -66.47916667 North Bounding Coordinate: 49.9375 South Bounding Coordinate: 24.0625

3. Data Origin and Methods

3.1. Introduction

The NASA Earth Exchange (NEX) U.S. Downscaled Climate Projections (NEX US-DCP30) dataset is comprised of downscaled climate scenarios for the conterminous United States 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 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 33 models and scenarios that were produced and distributed under

CMIP5, as well as ensemble statistics calculated for each RCP from all model runs available. The purpose of these datasets is to provide a set of 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 10^2 km), 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 US-DCP30 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]. 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 complete dataset of downscaled CMIP5 climate projections to facilitate the assessment of climate change impacts in the United States. The dataset compiles over 100 climate projections from 33 CMIP5 GCMs (Table 1) and four RCP scenarios (as available) for the period from 2006 to 2099, as well as the historical experiment for each model for the period from 1950-2005. Each of these climate projections is downscaled over the conterminous US at a spatial resolution of 30 arc-seconds (approximately 800 meters), resulting in a data archive size of more than 17 TB (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] or online at the URL http://gdo-dcp.ucllnl.org/downscaled_cmip_projections/dcpInterface.html#About.

3.2 Methods

3.2.1 Datasets

Climate Model Data: We compiled over 100 climate projections from the 33 CMIP5 GCM simulations (Table 1) across the four RCP scenarios. Each of the climate projections includes monthly averaged maximum temperature, minimum temperature, and precipitation for the periods from 1950 through 2005 ("Retrospective Run") and from 2006 to 2099 ("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 100+ 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 derived from meteorological station observations and created using the PRISM (Parameter-elevation Regressions on Independent Slopes Model) system developed at Oregon State University [Daly et al. 1994]. PRISM is a knowledge-based system that incorporates ground-based climate measurements, a digital elevation model, and expert analysis of complex climate phenomena (e.g., rain shadows) to produce continuously gridded estimates of climate variables. The PRISM data used in development of this dataset include monthly maximum temperature, monthly minimum temperature, and monthly total precipitation from 1950 to 2010 over the conterminous US at a spatial resolution of 30 arc-seconds (approximately 800m).

3.2.2 Data Pre-processing

Before applying the downscaling method, all data (GCM & PRISM) are interpolated to a common 1-degree grid. In addition, because the BCSD method does not explicitly adjust the trends (the slopes, in particular) in climate variables produced by GCMs, we extract the large-scale climate trends from the GCM 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 observationally-based PRISM data. This step is done at the scale of the common 1-degree grid to reduce computational costs. For each climate variable in a given month, the algorithm generates the cumulative distribution function (CDF) for the PRISM data and for the retrospective GCM simulations, respectively, by pooling and sorting the corresponding source values 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 observations. Based on this map, GCM values in any CDF quantile (e.g., p=90%) can be translated to corresponding PRISM 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 PRISM CDF, and then accepts the latter as the adjusted climate predictions. The climate projections adjusted in this way have the same CDF as the PRISM 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 climate trends are added back to the adjusted GCM climate fields.

3.2.4 Spatial Disaggregation (SD)

The Spatial-Disaggregation step spatially interpolates the bias-corrected GCM data to the finer resolution grid of the 30-arc second PRISM 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 monthly climatologies of the PRISM variables (temperature and precipitation) are generated at both native and aggregated resolutions (30 arc-seconds and 1-degree, respectively). The climatology for the SD step is the average for each month of the year calculated over the reference period, 1950-2005. Second, for each time step, the algorithm compares the bias-corrected GCM variables with the corresponding PRISM climatology to calculate "scaling factors". In particular, the scaling factors are calculated as the differences between the -corrected GCM and the PRISM 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 PRISM grid. Finally, the scaling factors are applied, by addition for temperatures and by multiplication for precipitation, on the fine-resolution PRISM climatologies 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 neighborhoods and communities. 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 2099, 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 PRISM 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 2006-2099 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.

4.4 Interannual Variability in Ensemble Statistics

This dataset includes a set of ensemble statistics for each RCP, including the ensemble mean, median, and 25th and 75th percentiles. The ensemble means are useful representations of the overall predicted change included in the entire ensemble of GCM runs for each RCP. The ensemble means may not be suitable for analyses that are likely to be sensitive to intra- and inter-annual variability in temperature and precipitation, however, since use of the ensemble means inherently compresses the month-to-month variability contained in each individual CMIP5 scenario. This is due to the fact that the ensemble member GCM runs are independent, and while use of the ensemble means may be helpful in reducing any residual biases, the high and low values in the ensemble at each step cancel each other, dampening the monthly and yearly variability that is present within each ensemble member.

5. Credits and Acknowledgements

Please cite this dataset as:

Thrasher, B., J. Xiong, W. Wang, F. Melton, A. Michaelis and R. Nemani (2013), Downscaled Climate Projections Suitable for Resource Management, Eos Trans. AGU, 94(37), 321.

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

> Climate scenarios used were from the NEX-DCP30 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.

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., and R. Nemani, 2013. New downscaled climate projections suitable for resource management in the U.S. *Eos, Transactions American Geophysical Union, 94(37), 321.*

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 0 – 26 April 2013 – Document created. This is a new document/dataset.

Rev 0.1 – 10 September 2013 – Update Appendix I and dataset citation note.

Rev 0.2 – 17 September 2013 – Updated Appendix I to include Known Issues / CDO *sellonlatbox* error.

Table 1. CMIP5 models included in downscaled archive

ACCESS1-0	FIO-ESM	IPSL-CM5A-MR
BCC-CSM1-1	GFDL-CM3	IPSL-CM5B-LR
BCC-CSM1-1-M	GFDL-ESM2G	MIROC-ESM
BNU-ESM	GFDL-ESM2M	MIROC-ESM-CHEM
CanESM2	GISS-E2-H-CC	MIROC5
CCSM4	GISS-E2-R	MPI-ESM-LR
CESM1-BGC	GISS-E2-R-CC	MPI-ESM-MR
CESM1-CAM5	HadGEM2-AO	MRI-CGCM3
CMCC-CM	HadGEM2-CC	NorESM1-M
CNRM-CM5	HadGEM2-ES	
CSIRO-MK3-6-0	INMCM4	
FGOALS-g2	IPSL-CM5A-LR	

APPENDIX I – WORKING WITH THE NETCDF FILES

The NEX-DCP30 native file format is in Network Common Data Form (netCDF). If one plans to work with the netCDF files directly, then a set of libraries and utilities that can read netCDF files must be installed on the compute platform you are working on. The netCDF-C libraries and utilities are available from the University Corporation for Atmospheric Research (UCAR) Unidata's program website: https://www.unidata.ucar.edu/software/netcdf/ and https://www.unidata.ucar.edu/downloads/netcdf/index.jsp . Prebuilt binaries may be downloaded and quickly installed, or there are a set of build from source and install instructions available; see the respective webpages.

Note: In this short guide, we use '%' to indicate a terminal prompt, and '>>>' to indicate a python interpreter prompt.

Once the base netCDF-C libraries and utilities have been installed, one can use the *ncdump* command line program to conveniently view metadata information contained within a netCDF file. Below is a command line invocation of *ncdump* from a terminal, "dumping" the file header information or metadata of a single data file (in this case, the example file is "tasmin_quartile25_amon_historical_CONUS_195001-195412.nc"):

% ncdump **-h** tasmin_quartile25_amon_historical_CONUS_195001-195412.nc

The output of the pervious command should include:

```
netcdf tasmin_quartile25_amon_historical_CONUS_195001-195412 {
dimensions:
        time = UNLIMITED ; // (60 currently)
        lat = 3105 ;
        lon = 7025 ;
        bnds = 2 ;
variables:
        float tasmin(time, lat, lon) ;
                tasmin:_FillValue = 1.e+20f ;
                tasmin:associated_files = "baseURL: http://cmip-
pcmdi.llnl.gov/CMIP5/dataLocation gridspecFile: gridspec_atmos_fx_BNU-
ESM_rcp26_r0i0p0.nc areacella: areacella_fx_BNU-ESM_rcp26_r0i0p0.nc" ;
.
.
.
}
```

Tip: It's important to include the **–h** option with the *ncdump* command if you whish to view the metadata only, otherwise all the data contained in the file will be printed to the terminal screen. See the *ncdump* documentation for more details on using this command.

A convenient way to access the netCDF files using the python program language (http://www.python.org) is available. A python extension to the netCDF-C library, netCDF4 (https://code.google.com/p/netcdf4-python/), provides a number of useful capabilities for working with the native data files. To illustrate the python netCDF4 module usage, we provide a simple example below. Note that lines starting with a '#' are comments within this snippet (see https://wiki.python.org/moin/SimplePrograms). Frist, we invoke the *python* interpreter from a terminal:

```
% python

Python 2.7.3 (default, Aug 13 2013, 15:59:28)

[GCC 4.3.4 [gcc-4_3-branch revision 152973]] on linux2

Type "help", "copyright", "credits" or "license" for more information.

>>> # Import the required modules

>>> import netCDF4

>>> import numpy

>>> # To open a netCDF data file in read only mode, note the "r"

>>> dataset = netCDF4.Dataset("tasmin_quartile25_amon_historical_CONUS_195001-
195412.nc", "r")

>>> # Printing the "variables" attribute will list the variables that are in the data file…

>>> print dataset.variables

OrderedDict([(u'tasmin', <netCDF4.Variable object at 0xb2bcd0>), (u'time', <netCDF4.Variable
object at 0xb2bc50>), (u'lat', <netCDF4.Variable object at 0xb2b1d0>), (u'lon',
<netCDF4.Variable object at 0xb2b3d0>), (u'time_bnds', <netCDF4.Variable object at
0xb2b5d0>), (u'lat_bnds', <netCDF4.Variable object at 0xb2bd50>), (u'lon_bnds',
<netCDF4.Variable object at 0xb2bdd0>)])

# A simple way to determine the dimensions of a variable

>>> print dataset.variables["tasmin"].shape

(60, 3105, 7025)

>>> # The above shows the tasmin variable has 60 "time steps" and each time step has a grid of
3105 rows (lat) and 7025 columns (lon). See ncdump example above.

>>> # Here, we retrieve first time step from the tasmin variable (all 3105x7025 items of  time
step 0)
```

```
>>> tasminS0 = dataset.variables["tasmin"][0]

# To determine what value was used as a "fill value" for the variable

>>> print dataset.variables["tasmin"]._FillValue

1e+20

>>> # To determine the minimum and maximum value

>>> tasminS0.min()

246.70547

>>> tasminS0.max()

290.28113

>>> # To extract a spatial subset of our first time step

>>> tasminS0Sub = tasminS0[10:50,20:30]

>>> # Finally, to close the netCDF file descriptor…

>>> dataset.close()
```

There are several other tools available that have data manipulation, interrogation and summarization capabilities. The Climate Data Operators (cdo, https://code.zmaw.de/projects/cdo/) command line program may be useful. An example invocation of *cdo* from a terminal:

```
% cdo runmin,60 tasmin_quartile25_amon_historical_CONUS_195001-195412.nc out.nc
```

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/).

See the cdo documentation for details of the many operators provided by the tool.

Potential Problems and Known Issues:

Note, some cdo operations such as *sellonlatbox:*

```
% cdo sellonlatbox,-90,-55,0,360 60 tasmin_quartile25_amon_historical_CONUS_195001-195412.nc out.nc
```

may fail due to an quirk with the way cdo expects an attribute 'time:bounds' for the variable 'time'. The issue with the 'time:bounds' attribute is being resolved (as of September 17, 2013) so that those who would like to use certain cdo operations may do so without any issues or limitations.