Geospatial Analysis Tool Kit for Regional Climate Datasets (GATOR)

An Open-source Tool to Compute Climate Statistic GIS Layers from Argonne Climate Modeling Results

James Kuiper, Veerabhadra R. Kotamarthi, Andrew Orr, and Jiali Wang
Argonne National Laboratory
Environmental Science Division
Lemont, Illinois

August 2017

Contents

Acknowledgments

Funding for this research provided by the Strategic Environmental Research and Development Program (SERDP) under contract RC-2242. This research used resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. This research used resources of the Argonne Leadership Computing Facility at Argonne National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under contract DE-AC02-06CH11357.

Introduction

This is a user manual for the geospatial analysis tool kit developed by Argonne National Laboratory to access and analyze regional-scale climate change datasets produced as a part of a project funded by SERDP. This document serves as a companion to the two other reports produced for the SERDP-funded project RC-2242, "Climate Change Impacts at the Department of Defense Installations." The two reports generated for the project are (1) *Climate Change Impacts to Defense installation: Final Report* (Kotamarthi et al., 2016b) and (2) *User Manual for Climate Model Outputs* (Kotamarthi et al., 2016a).

The tool kit described herein serves as an access point for a selected portion of the nearly 600 TB of data produced for the project in a format that is more familiar to users in the U.S. Department of Defense community. This large repository of climate model results for North America (Wang and Kotamarthi 2013, 2014, 2015) is stored in Network Common Data Form (NetCDF) (Unidata 2017). This report includes:

1. User documentation for GATOR.py, an open-source Python tool that uses Environmental Systems Research Institute (ESRI) ArcGIS to compute climate statistics from climate model results in NetCDF format and output them as GIS layers;

2. A description of the repository of condensed climate data for North America containing five variables (daily precipitation, relative humidity, temperature, solar radiation, and surface wind) generated by the Weather Research and Forecasting (WRF) model, version 3.3.1, driven by boundary conditions obtained from six independent global-scale climate models and two different future greenhouse gas emission scenarios, for three 10-year time periods (1995 to 2004, 2045 to 2054, and 2085 to 2094); and

3. Results and data for selected climate statistics computed from the model results and saved in Geographic Information System (GIS) format.

GATOR facilitates computing climate statistics for a set of 12 selected Department of Defense locations (Figure 1). This report and the GIS database provide examples of climate statistics that are useful for infrastructure planning, and studies of the potential effects of climate change. GATOR can compute many additional statistics by using new combinations of the existing input parameters. The code is open source and designed to be easily extendable to accommodate new input datasets, variables, or statistic types.

Figure 1. Selected Department of Defense Locations Configured in GATOR.

User Documentation for GATOR

GATOR is a software tool that was designed to compute climate statistics by processing the condensed climate model results provided with this report, and to output GIS layers with the results. It was written in Python 2.7 and uses ESRI ArcGIS 10.3. The GIS data repository provided with this report was created using GATOR with Python 2.7.8 and ESRI ArcGIS 10.3.0.4322 on a Windows 7 operating system; therefore, it has been most extensively tested with those versions (the source code is provided[1]). GATOR outputs are stored in ESRI file geodatabases, which have an open-source application programming interface (available at http://appsforms.esri.com/products/download/#File_Geodatabase_API_1.3), and most open-source GIS tools support the ESRI file geodatabase format.

GATOR employs a command-line design (see below for details) because some runs can take hours to complete. Also, sets of many runs can easily be initiated from a batch file listing them, and logs of the text output can easily be captured in files for later checking.

System Requirements

- Windows 7 or higher.
- 850 GB of free disk space if downloading or copying all project files, or at least 10 GB if using a version provided on external disk.

---

[1] The source code for GATOR is designed to be relatively easy to understand and modify, including systematic dictionaries for input defaults, file details, statistic types, output naming conventions, etc., and comments explaining key sections. It uses ESRI ArcGIS 10.3+, and is not compatible with recently released ESRI ArcGIS Pro 2.0, mainly due to its use of Python 3.x which has syntax differences from Python 2.x. However, it could be ported to use ArcGIS Pro without difficulty. See http://pro.arcgis.com/en/pro-app/arcpy/get-started/python-migration-for-arcgis-pro.htm for more details.

- At least 4.00 GB RAM (more is recommended).
- ESRI ArcGIS for Desktop 10.3 Basic (or more recent 10.x version).
- See http://desktop.arcgis.com/en/arcmap/10.3/get-started/system-requirements/arcgis-desktop-system-requirements.htm for more details.

Installing GATOR

1. Install and authorize the license for ESRI ArcGIS 10.3 (or more recent 10.x version), making sure to install Python 2.7. Administrative privileges are required for this step. See http://desktop.arcgis.com/en/arcmap/10.3/get-started/quick-start-guides/arcgis-desktop-quick-start-guide.htm for more details.
2. The files for the project can be provided in several ways:
   a. We will copy the files to an external drive if you provide one. Contact jkuiper@anl.gov for details. If you received an external drive with the project files, the fastest performance is achieved if all files (totaling 850 GB) are copied to a local drive. However, they can also be copied to a network drive, or directly used from the external drive.
   b. The files can be downloaded from https://anl.box.com/v/GATOR.
3. In Windows Explorer, determine (or assign) the drive letter and path to the project files. These instructions, and the default setting for GATOR, assume the top-level directory containing the files to be "V:\GATOR".
4. Use the Windows Start menu to run a command prompt. Either
   a. Run Start → Accessories → Command Prompt, or
   b. Choose Start, type "`cmd`" in the search box, and click on cmd.exe to run it.
5. In the command window:
   a. Type "`V:`", or the drive letter noted in Step 3.
   b. Type "`cd GATOR\Python`".
   c. Type "`python`".
      i. If Python is found, it should print the version and a >>> prompt. Type "`import arcpy`". After a pause, if you get an error, continue with step 6. If another >>> appears without an error, type "`exit()`" to exit Python and continue with step 7.
      ii. If Python is not found and you get an error, continue with step 6.
6. Check to see that Python 2.7x was installed with ArcGIS 10.3+. The default location for the files is C:\Python27\ArcGIS10.3 for 32-bit, and C:\Python27\ArcGISx6410.3 for 64-bit.

a. If the files are present, type "`set PATH= C:\Python27\ArcGIS10.3;%PATH%`" or "`set PATH= C:\Python27\ArcGISx6410.3;%PATH%`" in the command window. Repeat Step 5c to test for Python and the presence of the arcpy library.

b. If the files are not present, update the ArcGIS installation from Step 1 to add Python.

7. Type "`python GATOR.py -h`" in the command window. A GATOR header should print with a list of usage information and details about the optional command-line arguments.

8. (Optional) If you will be using GATOR with a location of the files different from the default of "V:\GATOR", you can change the default in the Python code to match your configuration. First make a backup copy of the …\Python\tools\defaults.py file, then edit it with any Python or text editor (Idle, Notepad, etc.). Change the path on line 7 to match your configuration, taking care to follow the syntax of the existing code.

9. If Step 7 does not work properly, or you have any questions, please contact Jim Kuiper at 630-252-6206 or jkuiper@anl.gov.

Running GATOR

Setting up to run GATOR is similar to some of the installation steps above.

1. Use the Windows Start menu to run a command prompt. Either

    a. Run Start → Accessories → Command Prompt, or

    b. Choose Start, type "`cmd`" in the search box, and click on cmd.exe to run it.

2. In the command window:

    a. Type "`V:`", or the drive letter noted in Step 3 of the Installation instructions.

    b. Type "`cd GATOR\Python`".

    c. If Step 6a was needed during installation, type
    "`set PATH= C:\Python27\ArcGIS10.3;%PATH%`" or
    "`set PATH= C:\Python27\ArcGISx6410.3;%PATH%`"

3. The command window is now ready to run GATOR. (If desired, multiple command windows can be set up to run more instances at the same time.)

4. (Optional) Type "`python GATOR.py -h`" in the command window to verify that GATOR is working.

5. Type "`python GATOR.py`" with a combination of input parameters specifying what you would like to compute. Details about each parameter are explained below, followed by example runs with combinations of the parameters. The Results section also lists the command-line parameters used to run the provided results.

Each command-line argument is specified as an argument-value pair, and the pairs can be listed in any order. Table 1 lists the arguments, their default values, and descriptions. If you do not specify one of the command-line arguments, GATOR will use the default. The `-inpath` and `-outpath` parameters are only required if the path to the input NetCDF files, or the desired location for the output files, is different from the default listed in Table 1.

Table 2 lists additional details for the GATOR `-model` parameter that specifies which of the six climate model files to process (or whether climate observations should be used instead). Table 3 lists the choices for the `-scenario` parameter, which sets the climate model scenario and the time period to be used. GATOR will process all years for the selected scenario, unless they are changed using `-startyear` or `-endyear`. For example, `-scenario rcp45mid -startyear 2047 -endyear -2047` would be used to compute only year 2047 for the RCP4.5 scenario.

Table 4 lists the `-stattype` choices used to choose the type of statistic computed. GATOR uses the `-statvalue1` and `-statvalue2` parameters with some statistic types, as listed in the table, to specify thresholds and other details. Table 5 lists the available variables currently available in the input data repository and code. They are specified with the `-statvariable` parameter; some, like heat index and wind chill, are computed from combinations of the input values. Table 5 also lists the units used for GATOR parameters and output, some of which are converted from different units in the input files.

The `-site`, `-extent`, and `-buffer` parameters specify the geographic extent of the computations. To run GATOR for the grid cells overlapping one of the study sites, the entire grid extent, or the contiguous United States, use the `-site` parameter with one of the codes listed in Table 6. For any other geographic extent, use the `-extent` parameter with the full path to a point, line, or polygon shapefile or ArcGIS feature class. GATOR will compute results for the grid cells overlapping that area. The `-buffer` parameter followed by a numeric distance in miles can be used with either `-site` or `-extent` to extend the area by the specified distance.

Finally, specifying `-debug yes` for any run causes the raw NetCDF values used for processing to be saved to the output geodatabase as a series of tables, in addition to the output layer. We advise only using `-debug yes` for small geographic extents because the volume of NetCDF data output to these tables is large.

Table 1. GATOR Command-Line Arguments, Default Values, and Descriptions.

Argument	Default value	Description
-inpath	V:\GATOR\Data\netcdf\	Path to input data directory[1]
-outpath	.\Results\	Path to output directory[1,2]
-model	ncep	Climate model for inputs: [ccsm | ccsmbc | gfdl | gfdlnn | hadgem | ncep | observed][3]
-scenario	hist	Climate scenario and time period for inputs: [hist | rcp45mid | rcp45end| rcp85mid | rcp45end][3]
-startyear	(first available year for scenario)	Beginning year for statistics computations
-endyear	(last available year for scenario)	Ending year for statistics computations
-stattype	dygt	Statistic type: [dygt | dylt | gtwave | ltwave | gemaxconsec | lemaxconsec | monannave | monanntot | pcnt | pcnt_dygt | pcnt_dylt] [3]
-statvariable	tempmax	Statistic variable: [heatindex | precip | precipobs | relhum | solarrad | tempmax | tempmaxobs | tempmin | tempminobs| windchill | windspeed][3]
-statvalue1	(default depends on statvariable)	First parameter for statistic type

Argument	Default value	Description
-statvalue2	(default depends on statvariable)	Second parameter for statistic type
-extent	none	Full path to a GIS layer to be used for the extent to process; if they are also specified, -extent uses -buffer, and overrides -site
-buffer	0	Buffer (in miles) added to study site or extent layer for output extent
-site	us	Extent of area or study site to be computed: [all | us | apg | cannon | eglin | elm | drum | hood | lewis | mccoy | riley | stew | pend | ypg][3]
-debug	no	Saves unprocessed NetCDF data to output geodatabase for the extent analyzed

[1] Enclose in quotes if there are spaces in the path, and the ending \ is required.
[2] Indicates the directory from which GATOR is run. A full path can also be specified. The default value can be changed in tools/defaults.py to match the data location on the user's system, rather than specifying it for each run.
[3] Square brackets enclose the list of all the choices, separated by | delimiters.

Table 2. GATOR Model Parameter Codes, NetCDF Data Directory Names, Model Names, Time Periods, and Scenarios Included in Data Repository.

-model Code	NetCDF Data Directory Name	Model Name and Boundary Conditions	Time Period(s)	Historical/Scenarios
ccsm	wrf_ccsm_raw	CCSM4 (raw)	1995 to 2004 2085 to 2094	Historical, RCP 4.5/8.5
ccsmbc	wrf_ccsm_bc	CCSM4 (bias corrected)	1995 to 2004 2045 to 2054 2085 to 2094	Historical, RCP 4.5/8.5
gfdlnn	wrf_gfdl_no_nudging	GFDL ESM2G (bias corrected)	1995 to 2004 2045 to 2054 2085 to 2094	Historical, RCP 4.5/8.5
gfdl	wrf_gfdl_nudging	GFDL ESM2G (bias corrected)	1995 to 2004 2045 to 2054 2085 to 2094	Historical, RCP 4.5/8.5
hadgem	wrf_hadgem	HadGEM-ES (raw)	1995 to 2004 2045 to 2054 2085 to 2094	Historical, RCP 4.5/8.5
ncep	wrf_ncep-r2	NCEP-R2	1995 to 2004	Historical
observed	observed	Observed Measurements	1995 to 2004	Historical

Table 3. GATOR Scenario Type Codes, Descriptions, and Years in Provided NetCDF Data.

-scenario Code	Description	Years in Provided NetCDF Data
hist	Historical	1995 to 2004
rcp45mid	Representative Concentration Pathway 4.5 (emissions peak around 2040, then decline)	2045 to 2054
rcp45end	Representative Concentration Pathway 4.5	2085 to 2094
rcp85mid	Representative Concentration Pathway 8.5 (emissions continue to rise throughout the 21st century)	2045 to 2054
rcp85end	Representative Concentration Pathway 8.5	2085 to 2094

Table 4. GATOR Statistic Type Codes and Descriptions.

-stattype Code	First Parameter (-statvalue1)	Second Parameter (-statvalue2)	Description
dygt	Threshold value (e.g., use “90” for 90°F for temperature variables)	Ignored	Annual number of days greater than threshold value <p#1>.
dylt	Threshold value	Ignored	Annual number of days less than threshold value <p#1>.
gtwave	Threshold value	Length of period (e.g., use “3” for 3-day periods)	Annual number of <p2#-day> periods greater than threshold value <p#1>.
ltwave	Threshold value	Length of period	Annual number of <p2#-day> periods less than threshold value <p#1>.
gemaxconsec	Threshold value	Ignored	Annual longest sequence of consecutive days >= than threshold value <p#1>.
lemaxconsec	Threshold value	Ignored	Annual longest sequence of consecutive days <= than threshold value <p#1>.
monannave	Ignored	Ignored	Monthly and annual averages.
monanntot	Ignored	Ignored	Monthly and annual totals.
pcnt[1]	Percentile (1 to 99, default = 90)	Ignored	<p#1>th percentile. (Value for each cell that <p#1>% of values are below.)
pcnt_dygt	Percentile (1 to 99, default = 90)	Ignored	Computes baseline <p#1>th percentiles for “hist” years, then counts days greater than baseline values for all future scenarios.
pcnt_dylt	Percentile (1 to 99, default = 90)	Ignored	Computes baseline <p#1>th percentiles for “hist” years, then counts days less than baseline values for all future scenarios.
percentile[1]	Percentile (1 to 99, default = 90)	Ignored	<p#1>th percentile. (Value for each cell that <p#1>% of values are below.)

[1] “pcnt” and “percentile” compute the same results. “pcnt” uses less memory but takes longer to run, while “percentile” is faster but uses more memory, sometimes causing Python to crash if it exceeds the amount of addressable memory for larger analysis extents.

Table 5. GATOR Statistic Variable Codes and Descriptions.

-statvariable Code	Description	Units	Default Value
heatindex	Maximum daily heat index	Fahrenheit	90
precip	Total daily precipitation	Inches	0
precipobs[1]	Observed total daily precipitation	Inches	0
relhum	Mean daily relative humidity	Percent	80
solarrad	Mean daily solar radiation	Watts per square meter	300
tempmax	Maximum daily temperature	Fahrenheit	90
tempmaxobs[1]	Observed maximum daily temperature	Fahrenheit	90
tempmin	Minimum daily temperature	Fahrenheit	32
tempminobs[1]	Observed minimum daily temperature	Fahrenheit	32
windchill	Minimum daily wind chill	Fahrenheit	32
windspeed	Mean daily wind speed	Miles per hour	10

[1] “precipobs,” “tempmaxobs,” and “tempminobs” are only available with -model observed and -scenario hist, and the other variables are not available with -model observed and -scenario hist.

Table 6. GATOR Site Codes, Descriptions, and States.

-site Code	Description	State
all	Whole model grid[1]	All
us	Contiguous United States	Lower 48 states
apg	Aberdeen Proving Ground	Maryland
cannon	Cannon Air Force Base	New Mexico
eglin	Eglin Air Force Base	Florida
elm	Elmendorf Air Force Base	Alaska
drum	Fort Drum	New York
hood	Fort Hood	Texas
lewis	Fort Lewis	Washington
mccoy	Fort McCoy	Wisconsin
riley	Fort Riley	Kansas
stew	Fort Stewart	Georgia
pend	Camp Pendleton	California
ypg	Yuma Proving Ground	Arizona

[1] Grid extents for -model ccsm* and ncep differ from gfdl*, hadgem, and observed.

Example Runs of GATOR

The following examples illustrate ways to use various combinations of GATOR parameters for a variety of purposes. See the Results section for the parameters used for these example output statistics. The code checks the inputs before computations begin. Helpful messages are output before computations begin if there are errors in the parameters, or input files are not available for the specified model, scenario, year(s), or variable.

1. `python GATOR.py`

   Simply running GATOR.py with no parameters causes it to use all the defaults. It prints the run's progress on the screen, and uses the default settings to compute the number of days with a maximum temperature over 90°F for the NCEP-R2 model, for the historic years of 1995 to 2004, for the contiguous United States. It uses input files from the default input path, saves an ESRI file geodatabase called TempF_DyGt_90_ncep_hist_<date>-<time>.gdb to the default Results subdirectory below the current directory, and will contain the layer TempF_DyGt_90_ncep_hist. GATOR uses the default path of V:\GATOR\Data\netcdf\ to find the input NetCDF files.

2. ```
   python GATOR.py -inpath c:\GATOR\data\netcdf\ -outpath .\
   -model ccsm -scenario rcp45end -endyear 2085 -stattype dylt
   -statvariable tempmin -statvalue1 32 -site apg -debug yes > myrun.log
   ```

   In this example, -inpath is used to specify a non-default location of the input NetCDF files. The ending `\` is required. Specifying `-outpath .\` causes the output geodatabase to be saved in the current directory rather than the default Results subdirectory below the current directory. The results of the run are output to a TempF_DyLt_32_ccsm_rcp45end layer within a TempF_DyLt_32_ccsm_rcp45end_<date>-<time>.gdb ESRI file geodatabase. In this case, GATOR will use the CCSM4 raw model results for the end-of-century RCP 4.5 scenario. Of the available 2085 to 2094 10-year period, only results for 2085 will be computed. It will count the number of days with minimum temperatures below 32°F for grid cells overlapping Aberdeen Proving Ground in Maryland. In addition, `-debug yes` will result in the raw input NetCDF data (365 daily minimum temperatures per grid cell) to be saved in a series of tables in the output geodatabase. Finally, the standard DOS ">" function is used to save the printed output in the file "myrun.log"
   ```

rather than displaying it on the screen. Below is a listing of the GATOR output from this example.

```
Geospatial Analysis Tool Kit for Regional Climate Datasets (GATOR)
==================================================================
Starting time: 07/21/2017 14:37:09

   Input data path:         v:\GATOR\data\netcdf\

   Output data path:        .\

   NetCDF input data from:  WRF CCSM(raw); with spectral nudging; 1-year spin-up time;
                            historical, RCP 4.5/8.5

   Selected scenario:       'rcp45end' - Representative Concentration Pathway 4.5
                            (Emissions peak around 2040, then decline)

   Years to process:        2085

   Statistic type:          'dylt' - Annual number of days less than threshold value
                            <p#1>.

   Statistic variable:      'tempmin' - Minimum Daily Temperature in Fahrenheit

   First parameter (#1):    32 Fahrenheit

   Site (used for extent):  'apg' - Aberdeen Proving Ground, MD


   Output geodatabase:      .\TempF_DyLt_32_ccsm_rcp45end_20170721-1437.gdb

   Output layer:            .\TempF_DyLt_32_ccsm_rcp45end_20170721-
                            1437.gdb\TempF_DyLt_32_ccsm_rcp45end

   Input NetCDF extent:     Rows 184 to 187, Columns 499 to 501


   Processing model 'ccsm', scenario 'rcp45end', year 2085...

      Input NetCDF file:    v:\GATOR\data\netcdf\wrf_ccsm_raw\
                            rcp45end\tmin_daily_2085.nc

      Output results field: DyLt_TempF_32_2085

      Processing row 184 (25.0%)...

      Processing row 185 (50.0%)...

      Processing row 186 (75.0%)...

      Processing row 187 (100.0%)...


Ending time:  07/21/2017 14:37:47

Elapsed time: 0:00:37
```

3. `python GATOR.py -model observed -scenario hist -stattype gtwave -statvariable precipobs -statvalue1 1 -statvalue2 7 -site lewis -buffer 100`

   Here, more extreme levels of observed precipitation (`-model observed` is specified with `-statvariable precipobs`) are analyzed for grid cells up to 100 miles around Fort Lewis, Washington. GATOR counts the number of 7-day periods with at least 1 inch of rain every day for all the historical years (1995 to 2004). Statistics for each year are computed, in addition to the 10-year total.

4. `python GATOR.py -model gfdl -scenario hist -stattype monannave -statvariable windspeed -extent c:\mygisdata\states\wyoming.shp -buffer 200`

   In this run, GATOR computes monthly and annual average wind speeds for the state of Wyoming (using a “wyoming” shapefile stored in a c:\mygisdata\states directory) and an area of 200 miles around it. It takes the input data from the bias-corrected GFDL ESM2G model results, and the historical years of 1995 to 2004. This could be useful for studying the potential for wind energy generation.

5. `python GATOR.py -model hadgem -scenario rcp85mid -stattype monannave -statvariable solarrad -site ypg`

   Here, GATOR computes monthly and annual average solar radiation for Yuma Proving Ground, for 10 mid-century years, assuming a scenario where emissions continue to rise, using the HadGEM model results.

6. All the above commands could be placed in a batch file called “myruns.bat”, then run in sequence by typing “`myruns.bat`” to start the run.

As the examples indicate, GATOR can produce a wide variety of results when the input parameters are varied. The output geodatabase, layer, and field names are designed to be self-documenting. Each one is output with a unique name and date-/timestamp, so multiple instances can be run concurrently in the same location without conflicts. All the input settings are printed at the beginning of the run to document the settings, followed by a log of the run’s progress.

Repository of Condensed Climate Model Data

Results from the Weather Research and Forecasting (WRF) 3.3.1 regional climate model, driven by boundary conditions from six global climate models and datasets (Table 7), are included in the data repository for this study. Provided results include daily time steps for near-surface maximum and minimum air temperatures (at 2 meters), near-surface wind (at 10 meters), precipitation (including both solid and liquid), near-surface relative humidity (at 2 meters), and solar radiation.

Appendix A provides further details about climate models and related terminology. More information about the broader project of creating the climate data repository, and analysis of the results, is documented in Kotamarthi et al. (2016a, 2016b). Electronic copies of these documents are included with the project files for this report in the GATOR\Documents directory.

Table 7. GATOR Abbreviation, Boundary Conditions for Regional Climate Model, Spectral Nudging, Nudging Strength, Time Periods, Spin-up Time, and Scenarios Included in Study

Abbreviation Used for GATOR	Boundary Conditions for Regional Climate Model[1]	Spectral Nudging	Nudging Strength	Time Period(s)	Spin-up Time	Historical/ Scenarios
ccsm	Community Climate System Model Version 4 (CCSM4), raw data	Yes	3×10^{-5} s^{-1}	1995 to 2004 2085 to 2094	1 year	Historical, RCP 4.5/8.5
ccsmbc	Community Climate System Model Version 4 (CCSM4), bias corrected.	Yes	3×10^{-5} s^{-1}	1995 to 2004 2045 to 2054 2085 to 2094	1 year	Historical, RCP 4.5/8.5
gfdlnn	Geophysical Fluid Dynamics Laboratory Earth System Model Version 2G, bias corrected	No	N/A	1995 to 2004 2045 to 2054 2085 to 2094	1 year	Historical, RCP 4.5/8.5
gfdl	Geophysical Fluid Dynamics LaboratoryEarth System Model Version 2G, bias corrected	Yes	$3 \text{ x } 10^{-5}$ s^{-1}	1995 to 2004 2045 to 2054 2085 to 2094	1 year	Historical, RCP 4.5/8.5
hadgem	Hadley Global Environment Model 2 - Earth System, raw data	No	N/A	1995 to 2004 2045 to 2054 2085 to 2094	1 year	Historical, RCP 4.5/8.5
ncep	National Centers for Environmental Predictions - Department of Energy Reanalysis 2	Yes	3×10^{-4} s^{-1}	1995 to 2004	1 day	Historical

[1] For brevity in this document, and for the GATOR -model parameter, the global climate model used for boundary conditions in the WRF model is simply called a model.

The NetCDF file library consists of files for each combination of model, scenario, time period, and variable type, as summarized in Table 8. The climate model results total 1,418 files and about 800 GB, and the 30 observed meteorology files total about 13 GB. Aside from the exceptions listed below, each NetCDF file contains daily values for one year and one variable type, indexed by grid row and column.

The directory organization and file naming convention is as follows:

netcdf
- gator_grids.gdb: File geodatabase containing grid extent layers, and site layers.
- observed
  - hist
    - Files for each <variable>_daily_<year>.nc
- wrf_ccsm_bc
  - hist
    - Files for each <variable>_daily_<year>.nc
  - rcp45end
    - Files for each <variable>_daily_<year>.nc
  - rcp45mid

Files for each <variable>_daily_<year>.nc
rcp85end
Files for each <variable>_daily_<year>.nc
rcp85mid
Files for each <variable>_daily_<year>.nc
<same for other models>

GATOR uses premade grids stored in gator_grids.gdb for processing. All climate model and observation data are at a 12-kilometer resolution (grid cell spacing), but the grid origins and spatial extents vary (Figure 2a). CCSM4 and NCEP-R2 results share the same grid origin and extent, and the GFDL, HadGem, and observation data share a second grid origin and extent. The two grids are also offset from one another as shown in Figure 2b. Finally, the observation data are only available for grid cells in the contiguous United States and a small part of Canada.

The following NetCDF files contain both the latitude/longitude grid cell centers, and the associated grid row (south_north) and grid column (west_east) indices matching the data files. They are not required by GATOR, but are provided as a reference:

- For ccsm and ncep grids: netcdf\wrf_ccsm_raw\lat_long.nc
- For gfdl, hadgem, and observed grids: netcdf\wrf_gfdl_nudging\wrfout_d01_2053-02-01_00_00_00.nc

There are some variations among the models in the numbers of days per year, and whether leap years are present, as follows:

- Ccsm, ccsmbc, gfdl, and gfdlnn: All years have 365 days (February 29 is omitted for leap years).
- Ncep: Leap years (1996, 2000, and 2004) have 366 days; all others have 365.
- Hadgem: All years have 360 days (all months have 30 days).
- Observed: Leap years (1996, 2000, and 2004) have 366 days; all others have 365.

GATOR takes these variations into account when computing monthly statistics.

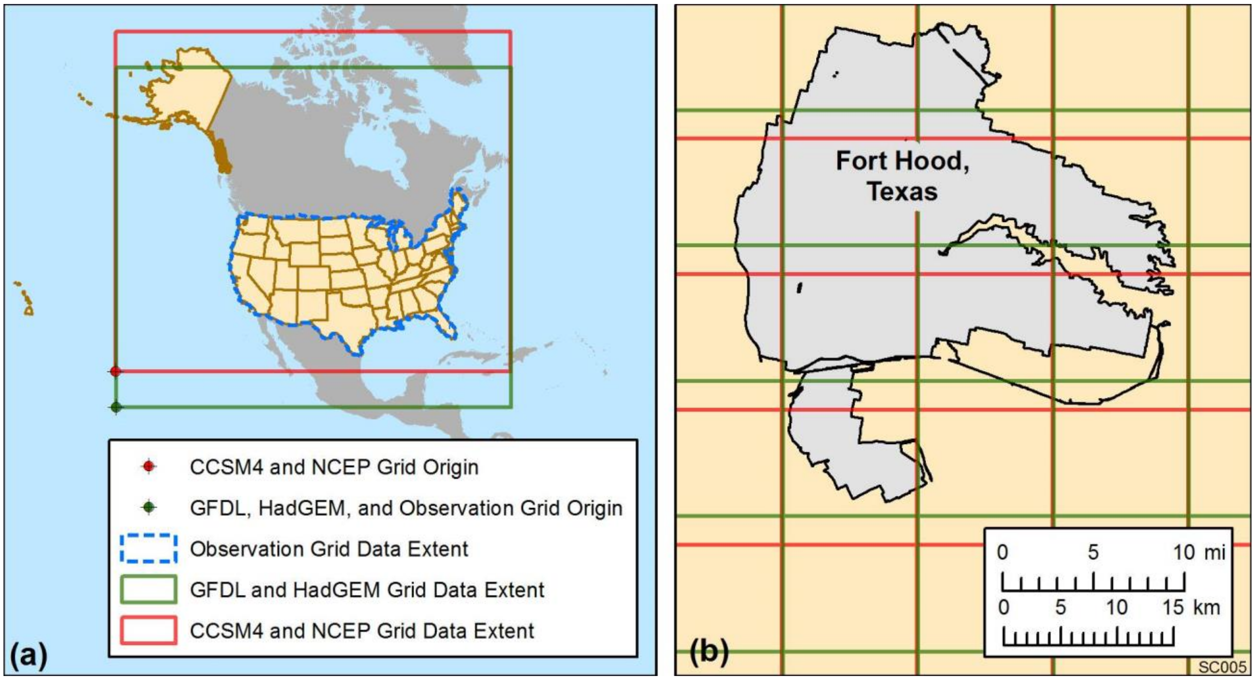

Figure 2. (a) Grid Origins and Data Extents of the Provided Climate Model and Observation Data, and (b) Large-Scale Example of the Grid Offset.

Table 8. Summary of NetCDF Files in Data Directories

Observed/Model	Scenario/Time Period				
	Historic	RCP 4.5		RCP 8.5	
	1995 to 2004	2045 to 2054	2085 to 2094	2045 to 2054	2085 to 2094
ccsm	All files	No files	All files	No files	All files
ccsmbc (bias corrected)	All files[1]	All except swd	All except rh, swd	All files	All files[2]
gfdl	All files	All files	All files	All files	All files
gfdlnn (no nudging)	All files[3]	All files	All files	All files	All files
hadgem	All files	All files	All files	All files	All files
ncep	All files	No files	No files	No files	No files
observed	All files	No files	No files	No files	No files

[1] Lacks relative humidity (rh) files for 2001 to 2004.
[2] Lacks relative humidity (rh) files for 2091 to 2094.
[3] Lacks solar radiation (swd) files for 2001 to 2004.

Table 5 lists the variables provided in the NetCDF data files with the units used in GATOR input parameters and output results. However, the data in the NetCDF files are sometimes in different units, or stored as more than one value. Table 9 lists the GATOR statistic variable code and variable description with the associated NetCDF variable name(s) and units, with notes about them. GATOR automatically converts the units. It also uses multiple fields or files to compute total precipitation, heat index, wind chill, and wind speed in the output units. All provided files use daily time steps, but they were generated from WRF model outputs with finer time steps. Means are based on 3-hour time steps, and maximum/minimum values are based on 40-second time steps. Days are represented as the sequential number of the day of the year, in UTC (Coordinated Universal Time), from 00UTC to the next 00UTC.

Table 9. GATOR Statistic Variable Codes, Variable Descriptions, NetCDF Variable Names, and NetCDF Units.

-statvariable Code	Variable Description	NetCDF Variable Name(s)	NetCDF Units
precip	Total daily precipitation	RAINC, RAINNC[1]	Millimeters
precipobs[2]	Observed total daily precipitation	PRCP	Millimeters
relhum	Mean daily relative humidity	rh2	Percent
solarrad	Mean daily solar radiation	SWDOWN	Watts per square meter
tempmax	Maximum daily temperature	T2MAX	Kelvin
tempmaxobs[2]	Observed maximum daily temperature	T2MAX	Kelvin
tempmin	Minimum daily temperature	T2MIN	Kelvin
tempminobs[2]	Observed minimum daily temperature	T2MIN	Kelvin
windspeed	Mean daily wind speed	U10, V10[3]	Meters per second

[1] RAINC represents precipitation (rain, snow, ice, graupel, etc.) from small-scale events that usually occur in the summer, while RAINNC is from large-scale events, usually in winter. They should be added together to determine total precipitation.

[2] "precipobs," "tempmaxobs," and "tempminobs" are only available with -model observed and -scenario hist, and the other variables are not available with -model observed and -scenario hist.

[3] U10 and V10 represent wind vectors (positive for west to east, and south to north, respectively). Computing the hypotenuse yields the wind speed.

The climate models perform computations and output results in the following projection:

Projection: Lambert Conformal Conic

Central Meridian:	-107.0
Standard Parallel 1:	60.0
Standard Parallel 2:	30.0
Scale Factor:	1.0
Latitude of Origin:	0.0
Units:	Meters

Datum: D Sphere EMEP

Spheroid:	Sphere EMEP, 6370000.0, 0.0
Prime Meridian:	Greenwich, 0.0
Units:	Degrees, 0.0174532925199433

Note that the models use a spherical mapping datum rather than the spheroidal 1984 World Geodetic System (WGS84) datum often used for mapping purposes. There is not an exact mathematical transformation between the EMEP sphere and WGS84, but the practice of the user community is to treat them as WGS84 when projecting model output data to more commonly used projections (Computing.io 2016; Krč 2015). The GIS database includes results both in the WRF climate model projection and in the more commonly used Albers equal area projection for the contiguous United States, and Alaska.

Results for Selected Climate Statistics

GATOR was used to compute eight sets of statistics as examples that will be useful for understanding potential future climate changes, and associated planning. Statistics for the historic period provide a baseline that is useful for comparing model results to historic observations, and to future projections. The following sections describe the statistics, some examples of their role in planning decisions, how they run in GATOR.py, and example results. Appendix B provides further details about the GIS database provided with this study, which contains the computed statistics as GIS layers.

Number of Days Over 90° Fahrenheit

This statistic is simply a count of the number of days per year, and per 10-year period, with a maximum temperature over 90° Fahrenheit. It is useful for understanding many potential effects:

- Limitations to outdoor activities;
- Requirements for air cooling systems and other infrastructure designs;
- Response activities to protect vulnerable populations;
- Peak load estimates for electricity consumption;
- Solar panel efficiency;
- Transmission line sag;
- Level of impact on power plant efficiency, and cooling water requirements;
- Temperature-related degradation of food, materials, and chemicals in the environment;
- Effects on domestic animals and wildlife; and
- Heat stress to crops and vegetation.

Figure 3 depicts results of this statistic for 2085 to 2094 from the Community Climate System Model, RCP 4.5.

This statistic is run in GATOR.py by using `–stattype dygt –statvariable tempmax –statvalue1 90`. (Similarly, cold temperature counts can be run by using `–stattype dylt –statvariable tempmin –statvalue1 32`.)

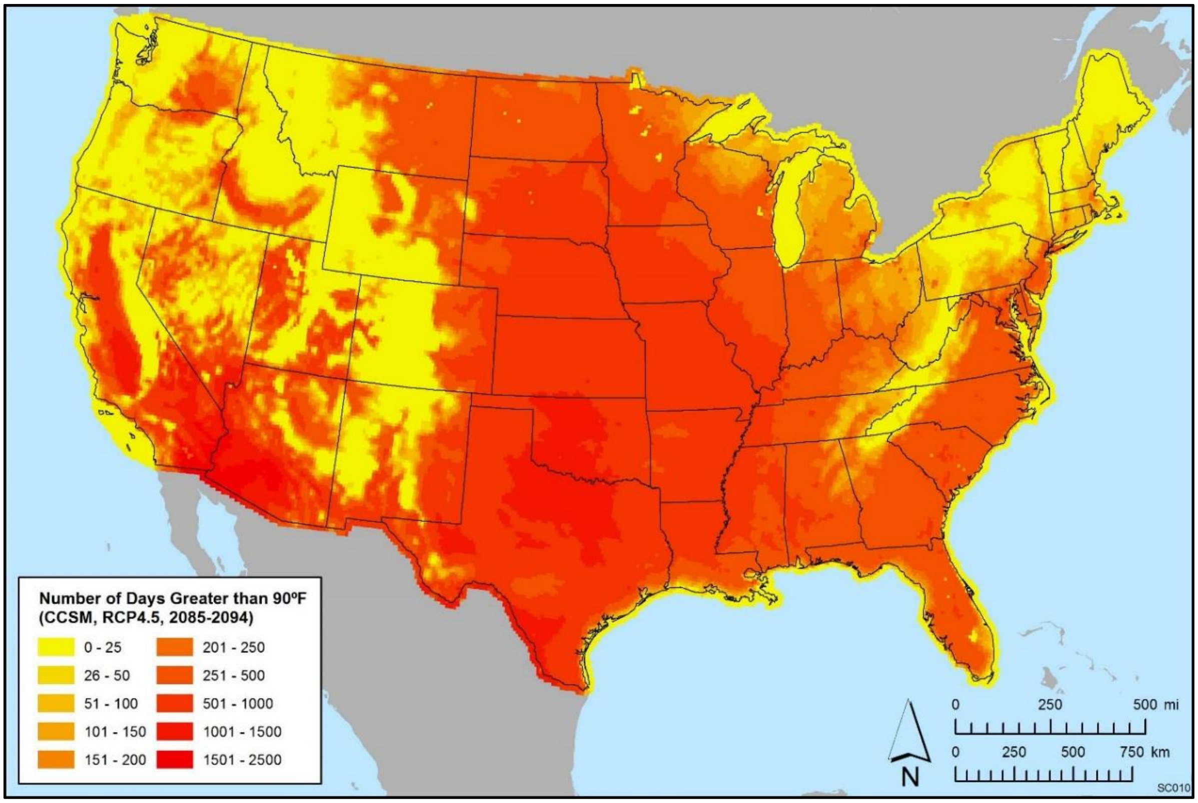

Figure 3. Number of Days with Near-Surface Air Temperature Exceeding 90°F from 2085 to 2094, under the RCP4.5 Scenario, Simulated by the WRF Model Driven by Raw CCSM4 Boundary Conditions.

Number of Days with a Heat Index Over 90° Fahrenheit

Heat index uses both temperature and humidity to measure what the temperature feels like to the human body *in a shaded location*. It can be 15°F higher in direct sunlight. Hot and humid conditions make it more difficult for the body to cool itself because less evaporation of sweat occurs. Heat index more precisely measures the level of comfort for outdoor activities and the risk of heat-related medical conditions than temperature alone.

Figure 4 depicts results of this statistic for 2085 to 2094 from the Community Climate System Model, RCP 4.5.

GATOR.py runs this statistic using `–stattype dygt –statvariable heatindex –statvalue1 90`. (Similarly, wind chill measures the effect of temperature and wind on cooling, and wind chill counts can be run using `–stattype dylt –statvariable windchill -statvalue1 32`.)

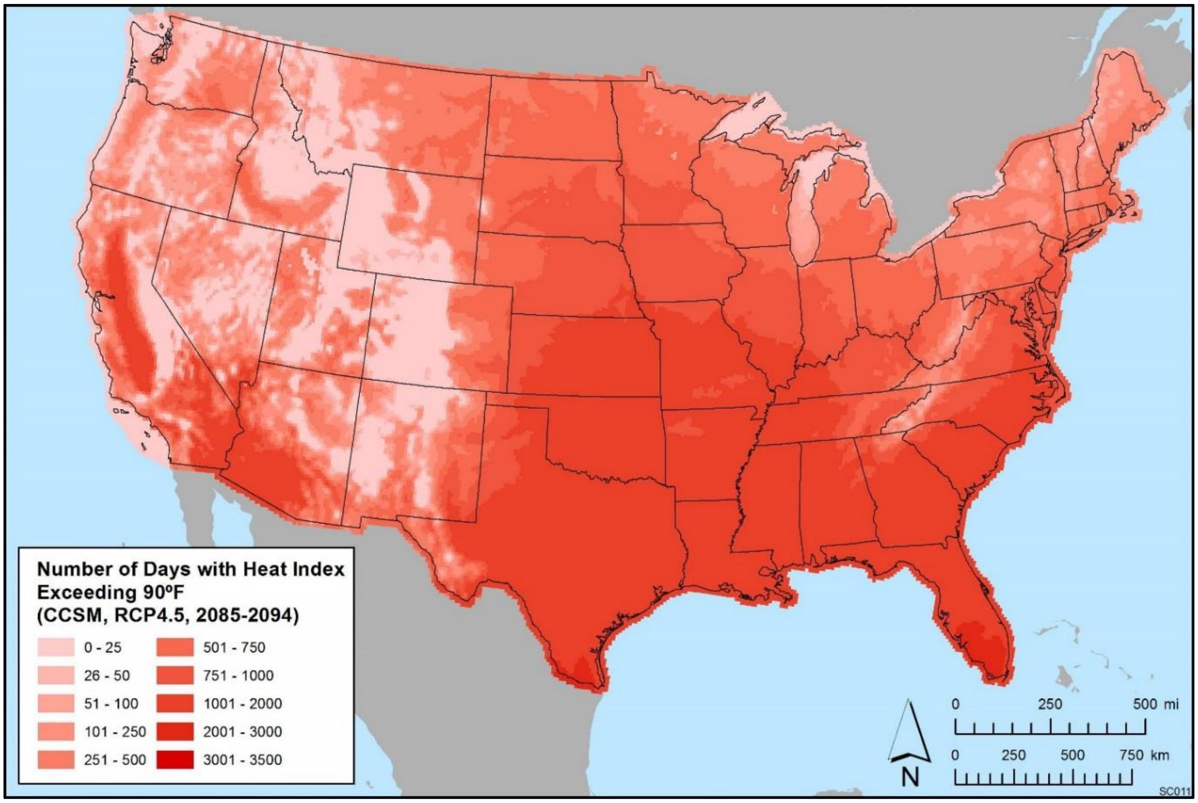

Figure 4. Number of Days with Near-Surface Heat Index Exceeding 90°F from 2085 to 2094, under the RCP4.5 Scenario, Simulated by the WRF Model Driven by Raw CCSM4 Boundary Conditions.

Heat Waves

Multi-day periods of high temperatures, or heat waves, were analyzed assuming a threshold of 90°F and a minimum length of 3 days. Heat waves have some unique considerations for planning that differ from the simple total number of days above a threshold temperature. For example:

- Power generation may be sufficient for a short peak in demand, but may not be able to sustain production for a longer period;
- Impacts on crops, vegetation, domestic animals, and wildlife may be greater during a sustained heat wave; and
- Medical services might be inadequate during sustained or more frequent heat waves.

Figure 5 depicts results of this statistic for 2085 to 2094 from the Community Climate System Model, RCP 4.5.

GATOR.py runs this statistic using `–stattype gtwave –statvariable maxtemp -statvalue1 90 –statvalue2 3`. (Similarly, cold snaps can be run using `–stattype ltwave -statvariable mintemp -statvalue1 32 –statvalue2 3`.)

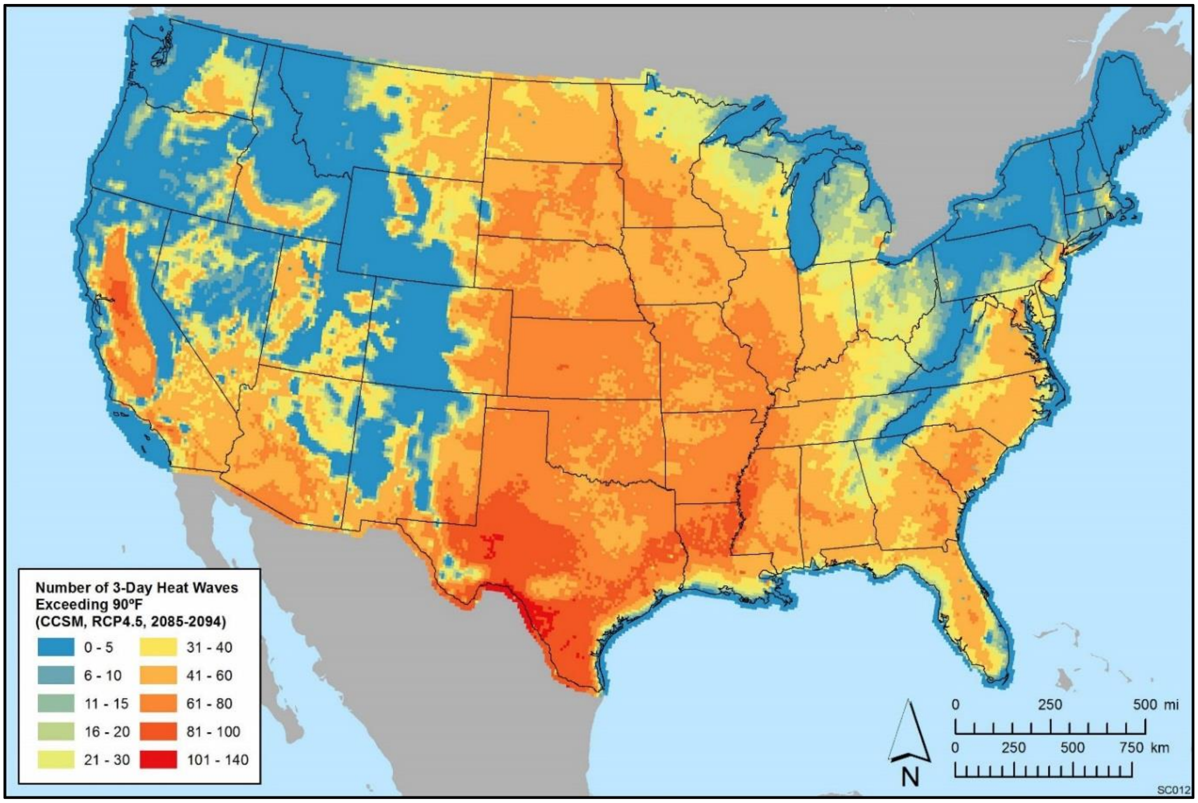

Figure 5. Number Three-day Heat Waves Exceeding 90°F from 2085 to 2094, under the RCP4.5 Scenario, Simulated by the WRF Model Driven by Raw CCSM4 Boundary Conditions.

Monthly and Annual Precipitation

These results provide total monthly and annual precipitation for each year, and average monthly and annual precipitation for each 10-year period. They provide several useful measures of precipitation, including variation in precipitation throughout a given year; variation in precipitation for the same month over different years; 10-year per-month averages to assess longer-term patterns between historical, midcentury, and end-of-century periods; and 10-year annual average precipitation. Mapping any of these values shows the spatial variation of precipitation.

These statistics can inform planning related to the following:

- Water supply for any type of consumption;
- Reservoir size and the amount of power generation or water supply it could provide;
- Whether climate change is expected to affect precipitation amounts, seasonal patterns, or spatial distribution of rainfall;
- Long-term viability of agricultural production;
- Design considerations for infrastructure resiliency;
- Potential for changes in wildfire frequency or intensity; and
- Impacts on sensitive ecosystems, vegetation, or wildlife.

Figure 6 depicts results of this statistic for 2085 to 2094 from the Community Climate System Model, RCP 4.5.

GATOR.py runs this statistic using `-stattype monanntot -statvariable precip`. (Similarly, temperature results can be generated with `-stattype monannave -statvariable maxtemp`, or wind speed results can be generated using `-stattype monannave -statvariable windspeed`.)

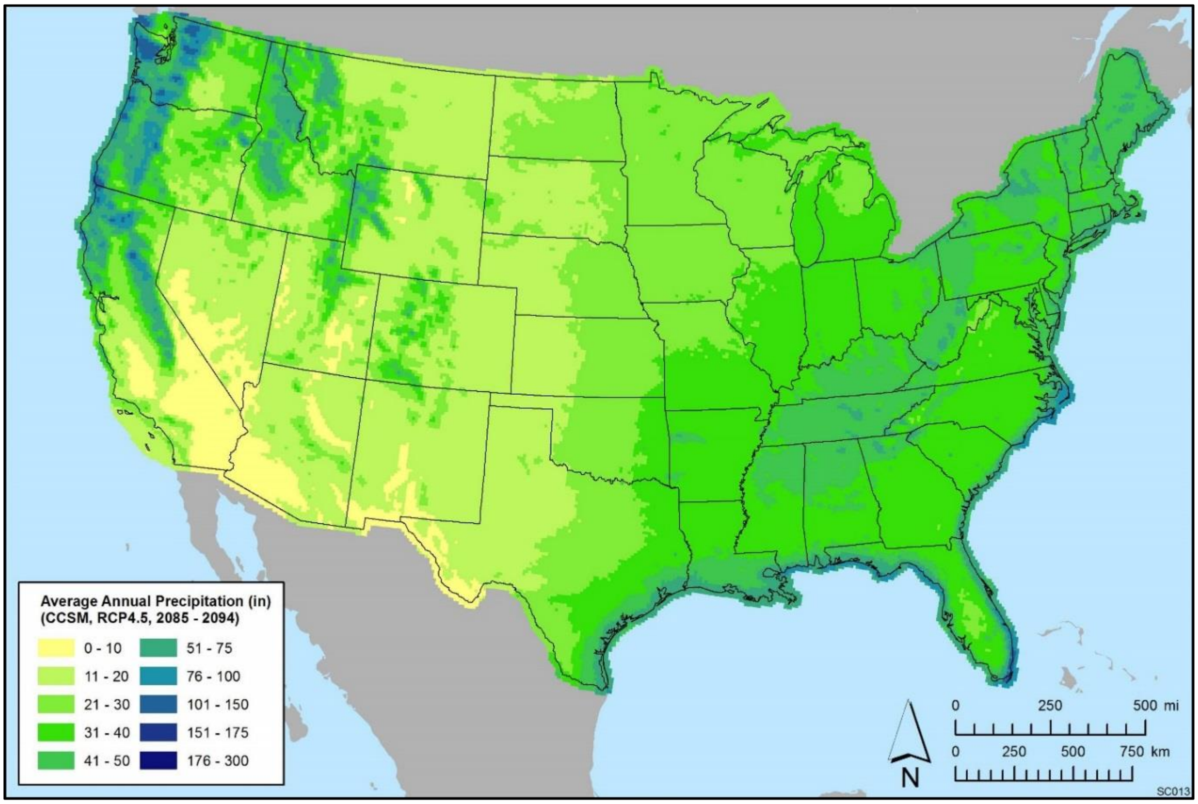

Figure 6. Average Annual Precipitation from 2085 to 2094, under the RCP4.5 Scenario, Simulated by the WRF Model Driven by Raw CCSM4 Boundary Conditions.

Drought

GATOR analyzed drought conditions by computing the largest number of consecutive days without precipitation. Investigating periods of low or no precipitation, and potential changes due to climate change is useful for planning in areas such as:

- Power plant cooling systems;
- Choice of power generation technology, and costs for cooling systems;
- Aquifer depletion or recharge;
- Viability of a region for different types of agriculture;
- Changes in natural vegetation and wildlife populations and distribution; and
- Frequency and intensity of wildfires.

Figure 7 depicts results of this statistic for 2085 to 2094 from the Community Climate System Model, RCP 4.5.

GATOR.py runs this statistic using `-stattype lemaxconsec -statvariable precip -statvalue1 0`. (Similarly, results related to permafrost or glacial melting could be obtained by using `-stattype lemaxconsec -statvariable maxtemp -statvariable 32`.)

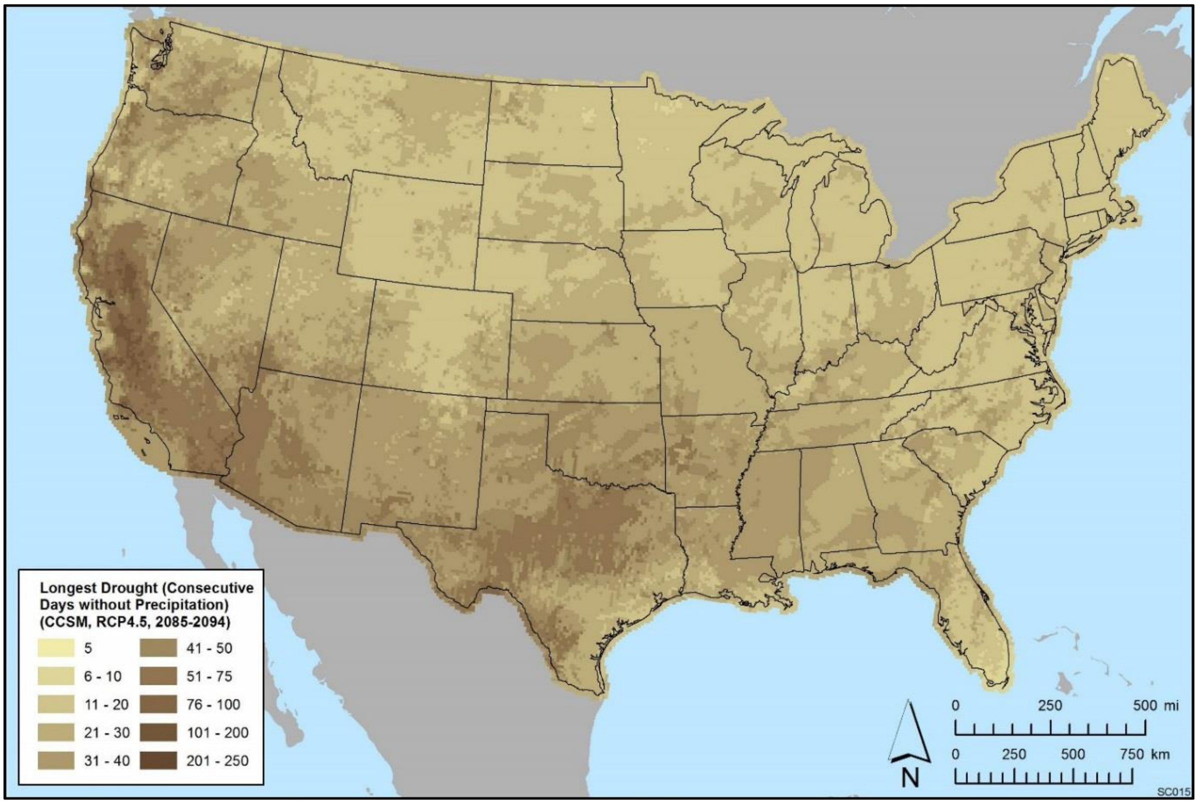

Figure 7. Longest Drought (Consecutive Days without Precipitation) from 2085 to 2094, under the RCP4.5 Scenario, Simulated by the WRF Model Driven by Raw CCSM4 Boundary Conditions.

Extreme (>90th Percentile) Precipitation

GATOR computed extreme precipitation levels using a 90th percentile statistic. The output values are the precipitation level at which 90% of the daily values are lower for a particular location. Compared to a static threshold throughout the geographic extent, percentiles are focused on the precipitation range for each cell and are more useful for comparing different scenarios or periods, allowing projected trends of increased or decreased rainfall to be mapped proportionally for both dry and wet regions.

This statistic is useful for planning in areas such as:

- Potential future trends in agricultural or natural vegetation viability;
- Power plant cooling systems;
- Choice of power generation technology, and costs for cooling systems;
- Aquifer depletion or recharge;
- Changes wildlife populations and distribution; and
- Frequency and intensity of wildfires.

Figure 8 depicts results of this statistic for 2085 to 2094 from the Community Climate System Model, RCP 4.5.

GATOR.py runs this statistic using `-stattype pcnt -statvariable precip -statvalue1 90`. (Similarly, levels of the highest wind speeds can be mapped using `-stattype pcnt -statvariable windspeed -statvalue1 90`, or levels of highest solar radiation can be mapped using `-stattype pcnt -statvariable solarrad -statvalue1 90`.)

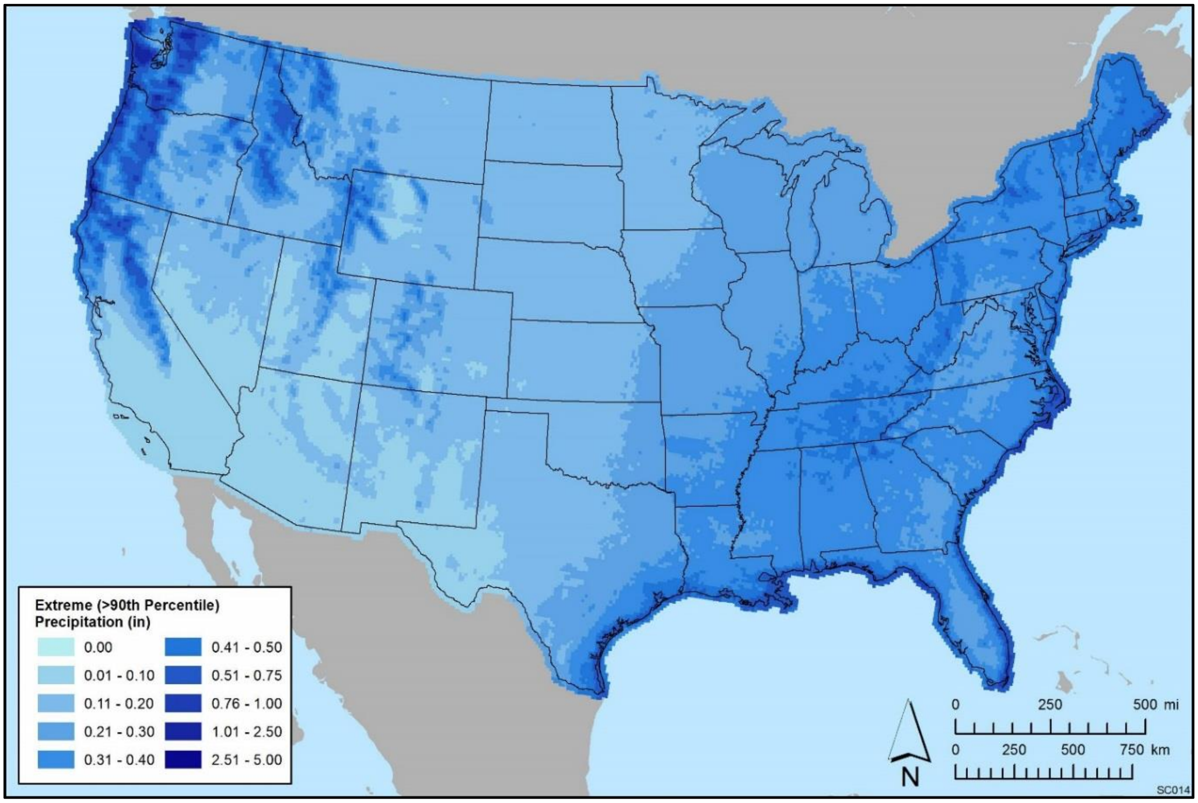

Figure 8. Extreme (Exceeding 90th Percentile) Precipitation, in Inches, from 2085 to 2094, under the RCP4.5 Scenario, Simulated by the WRF Model Driven by Raw CCSM4 Boundary Conditions.

Multi-model Averages

A second level of analysis included computing multi-model averages for each of the prior statistics, and each of the scenarios and periods. These average values provide a more generalized statistic that combines all the models in this study, and helps reduce the influence of one model's bias compared to the others. The results include up to five layers per statistic, including hist, rcp45mid, rcp85mid, rcp45end, and rcp85end, with up to six models averaged per layer. Some statistics have fewer models or scenarios/time periods, correlating to the available input NetCDF data summarized in Table 8.

This work was done by:

(1) Copying the compiled results geodatabase containing the above six statistics into a separate working geodatabase;

(2) Removing the per-year statistic fields, keeping only the 10-year results;

(3) Creating a unique identifier called "RowCol" for every cell (including a shift to the gfdl and hadgem grids to match the nearest cell in the ccsm and ncep grids);

(4) Renaming the 10-year fields to the model abbreviation (e.g., renaming the "DyGt_HtIdxF_90_1995_2004" field in the DyGt_HtIdxF_90_ccsm_hist layer to "ccsm" to prepare for the next step;

(5) Joining all model results for a particular statistic and scenario/time period into one layer;

(6) Adding a model average field and computing the average; and

(7) Removing unnecessary extra fields resulting from step 5.

This process was automated to run for all the layers using a series of Python programs. (These are not provided with the project files, because they were hard coded to the specific statistics and data layers above; however, they can be shared if needed.)

Figure 9 depicts results of this statistic for 2085 to 2094, averaging five models, for RCP 4.5.

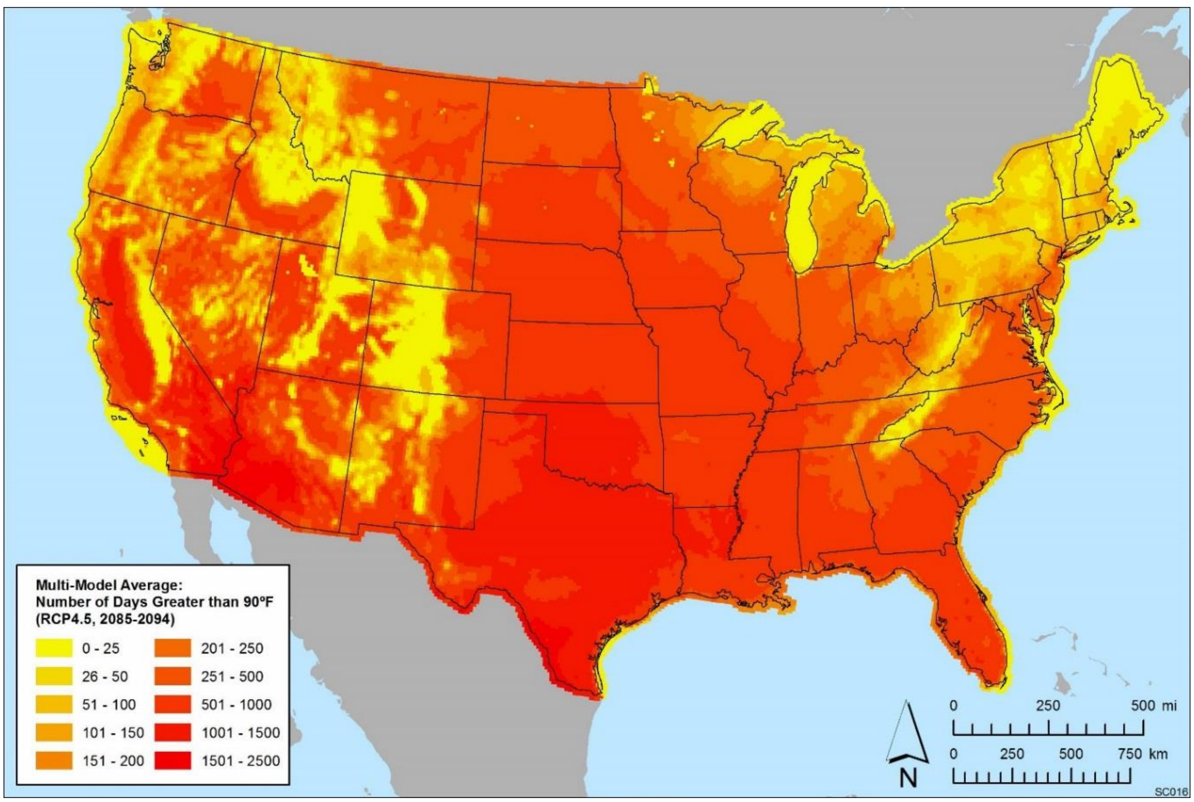

Figure 9. Average of Results from the WRF Model Driven by Raw CCSM4, Bias-Corrected CCSM4, GFDL ESM2G without Nudging, GFDL ESM2G with Nudging, and Raw HadGEM-ES Boundary Conditions for the Number of Days with Near-Surface Air Temperature Exceeding 90°F Statistic from 2085 to 2094, under the RCP4.5 Scenario.

Model Bias

Model bias can be estimated by comparing model results to observed values for the same period. For this statistic, only the multi-model averages for the 10-year statistics were computed for each of the statistic types, due to time and cost limitations, and the excessive number of combinations that would result for doing each individual model. In addition, Kotamarthi et al. (2016b) provides an extensive analysis of model bias. The absolute difference (model average – observed), and the percent difference ((model average – observed) / observed * 100) was computed using the following approach:

(1) Remove unnecessary extra fields from the model average layer for each input statistic, and scenario/time period except for the 10-year model average(s) and the RowCol unique ID;

(2) Add a "RowCol" unique ID to the layer with the matching statistic generated from historical meteorological observations;

(3) Join the observed values layer to the model average layer using the unique ID;

(4) Save it as an "ObsCompare" layer;

(5) Add "ObsDiff" and "ObsPcntDiff" fields and compute the values; and

(6) Remove unnecessary extra fields resulting from step 3.

This process was automated to run for all the layers using a Python program. (This program is not provided with the project files, because it was hard coded to the specific statistics and data layers above; however, it can be shared if needed.)

Figure 10 depicts the difference between the multi-model average and observed near-surface temperature for the number of days exceeding 90°F statistic, for 1995 to 2004. Figure 11 depicts the percent difference between the multi-model average and observed average annual precipitation, for 1995 to 2004.

Results of this statistic are useful for spatially mapping the direction and magnitude of the differences, or percent differences, in the model results from observed measurements, and help in understanding where and how well the models match observed patterns.

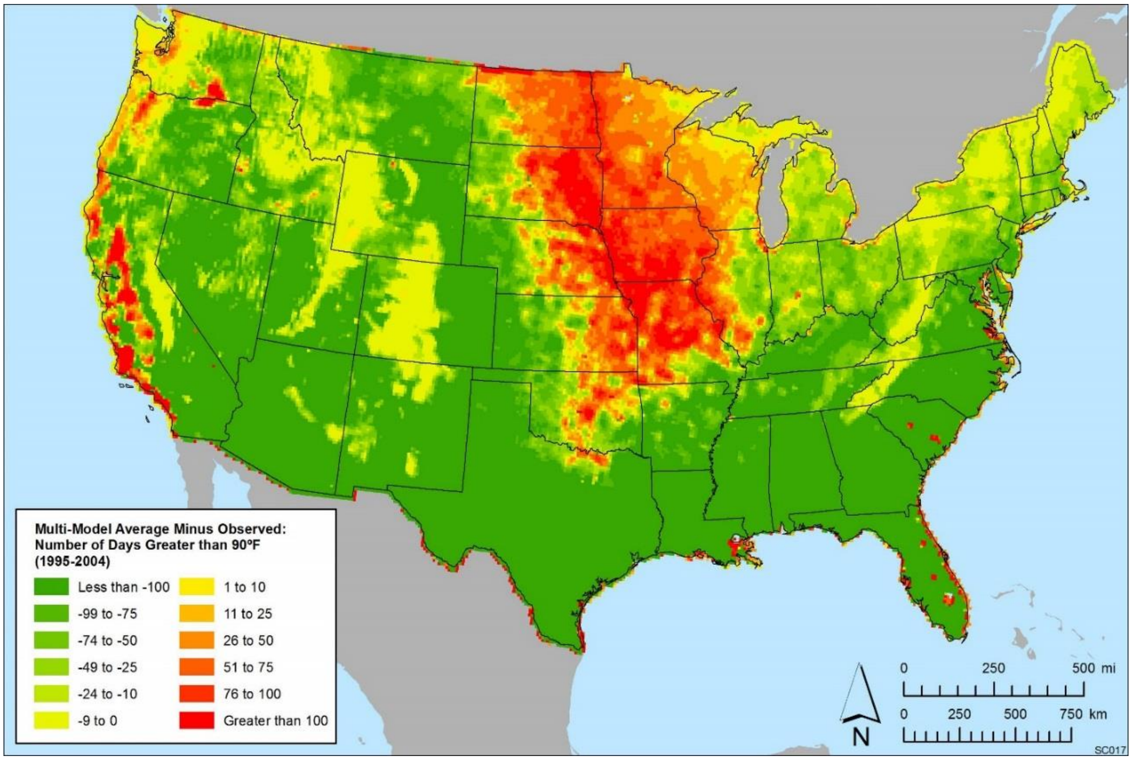

Figure 10. Difference between Multi-Model Average and Observed Number of Days with Near-Surface Air Temperature Exceeding 90°F from 1995 to 2004.

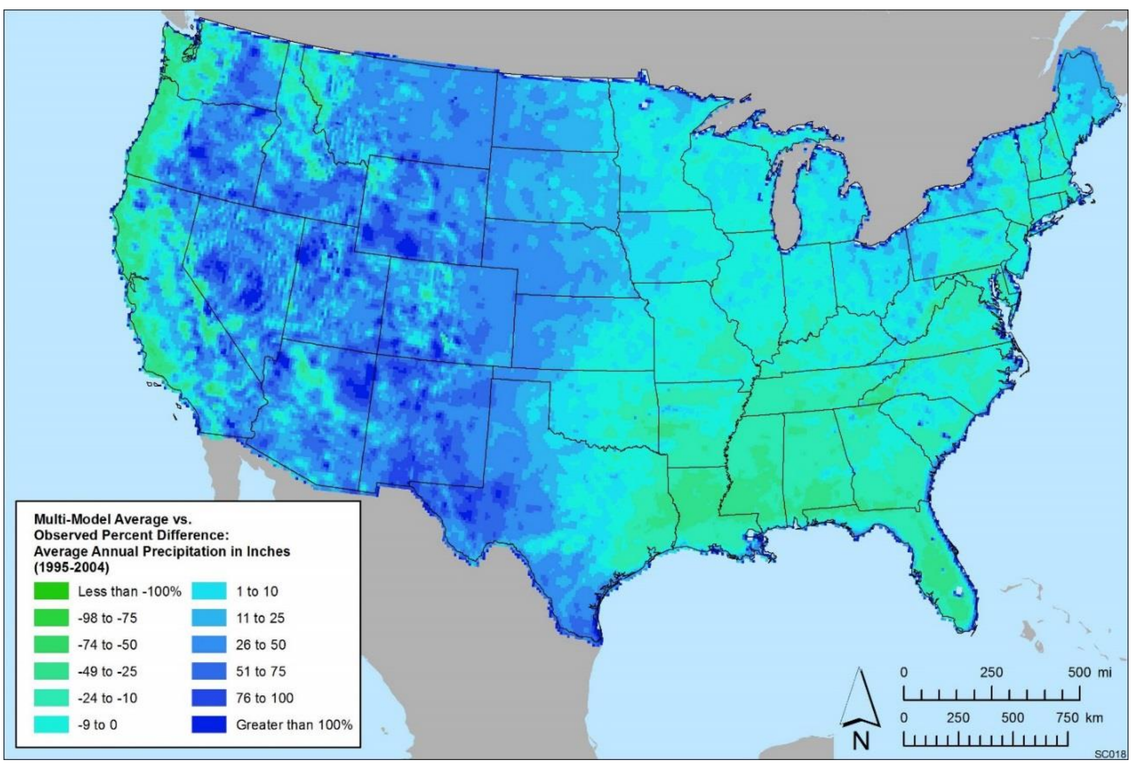

Figure 11. Percent Difference between Multi-Model Average and Observed Annual Average Precipitation from 1995 to 2004.

References

Computing.io (2016) *WRF: Latitudes, Ellipsoids, and Datums*. Available at: http://computing.io/wp/2016/01/wrf-latitudes-ellipsoids-and-datums. Accessed 6/20/2017.

Kotamarthi, R., L. Mearns, K. Hayhoe, C. Castro, and D. Wuebbles (2016a) *SERDP Report, Use of Climate Information for Decision-Making and Impacts Research: State of Our Understanding.* Prepared for the Department of Defense, Strategic Environmental Research and Development Program. 55pp.

Kotamarthi, R., J. Wang, Z. Zobel, D. Wuebbles, K. Hayhoe, M. Stein, and D. Changnon (2016b) *Climate Change Impacts at Department of Defense Installations, Resource Conservation and Climate Change (RC) Program Area, CRSON-12-02, Final Repor*t.

Krč, P. (2015) *The Pitfalls of Lambert Conformal Projection in WRF and MM5*. Available at: http://www.pkrc.net/wrf-lambert.html. Accessed 6/20/2017.

Unidata (2017) *Network Common Data Form* (*NetCDF)*. UCAR/Unidata Program Center, Boulder, CO. Available at: http://www.unidata.ucar.edu/software/netcdf. Accessed on 6/20/2017.

Wang, J., and V. R. Kotamarthi (2015) High-resolution dynamically downscaled projections of precipitation in the mid and late 21st century over North America, *Earth's Future*, 3, 268–288, doi:10.1002/2015EF000304.

Wang, J., and V. R. Kotamarthi (2014) Downscaling with a nested regional climate model in near-surface fields over the contiguous United States, *Journal of Geophysical Research, Atmosphere*, 119, 8778–8797, doi:10.1002/2014JD021696.

Wang, J., and V. R. Kotamarthi (2013) Assessment of Dynamical Downscaling in Near-Surface Fields with Different Spectral Nudging Approaches Using the Nested Regional Climate Model (NRCM), *Journal of Applied Meteorology and Climatology,* 52, 1576–1591.

Appendix A: Climate Modeling Background

Climate models, first developed in the early 1970s, use physical principals to represent the atmosphere at a global scale. They have continually improved over the past four decades. The latest versions of the models represent the atmosphere, ocean, and biosphere together and are referred to as earth system models (ESMs). These complex models represent hundreds of physical, chemical, and biological phenomena that occur at all times around the globe and determine the physical, chemical, and biological state of the atmosphere, ocean, and biosphere. The most recent evaluation of these global climate models (GCMs) included 40 different models (Flato et al. 2013). The evaluation concluded that these models predict changes in surface temperature over the recent past with high confidence. The current generation of models have improved surface temperature predictions and small improvements of precipitation prediction, compared to the previous generation of models. These models operate at a spatial grid resolution between 100 km^2 and 300 km^2. The respective model development teams continuously improve physical processes represented in these models. One major goal of the new model development is to obtain higher spatial resolution. In 5 to 10 years, GCMs with spatial resolutions of 25 km^2 or finer will likely be available (ACME 2014). Below is a discussion of some key characteristics and limitations of climate modeling.

Uncertainties

Climate model projections have three primary types of uncertainties: (1) scenario uncertainties, (2) model uncertainties, and (3) internal variability. The fraction contributed by the scenario uncertainty is the largest as one moves further away from the present. The model's internal variability causes the highest uncertainty at times closest to the present and the model uncertainty remains constant or slowly decreases throughout the entire projection period.

Scenario Uncertainty

Greenhouse gas (GHG) emission projections are generated for different socioeconomic scenarios. For consistency, the emissions scale to known emissions from the present and recent past. Therefore, near-term uncertainty from these emissions is not significant and does not contribute significantly to model projections of climate change close to the present day. However, the assumptions of the scenarios used to generate these emissions diverge from each other regarding fossil fuel use, technology, and other socioeconomic factors. As a result, the estimated emissions for each of these scenarios start diverging from each other after a few decades, and the divergence is significant by the end of the century. Climate modelers addressed this uncertainty by modeling a select number of emission scenarios that provide reasonable coverage of the emission uncertainty at the end of the century.

Model Uncertainty

Climate models represent many of the physical, chemical, and biological processes that make up the earth system and its interactions with incoming solar radiation. The models use fundamental principles of physics, chemistry, and ecology to represent known processes. However, a number of phenomena occur at spatial scales below those that are resolved in the current generation of climate models (for example, the formation of clouds). Several types of clouds are formed by physical processes that occur at very small spatial scales (hundreds of meters to kilometers), which are below the 100-km threshold of many current-generation climate models. Models represent these unresolved processes using a parametric approach. This introduces uncertainty, because the parametric models are only as good as the available observations that form the basis for developing the parameterizations. As we gather more detailed observations of such processes, the models representing them constantly improve. This is expected to reduce model uncertainty.

At present, the model uncertainty used in impact assessments can be represented in two ways: (1) generating a physical ensemble simulation with a single model to provide an uncertainty estimate for that model or (2) using results from multiple models that most likely have different parametric representations of a physical process. Although the former output may be more desirable for understanding the uncertainty in each model, this type of model is fairly expensive to produce, because the parameter range of each parametric model and the number of parametric models in a typical climate model can be fairly large. It is likely that the information from the latter will be more readily available to a user interested in impact assessments. Approximately 40 models were used in the most recent Intergovernmental Panel on Climate Change (IPCC) assessments and could be used to explore model uncertainty for impact assessments. However, most users pick a few models that represent the range of model responses to GHG emissions.

Internal Variability

Near-term climate projections are dominated by the internal variability of the models (Hawkins and Sutton 2009; Deser et al. 2012). GCMs are based on numerical equations and require the description of the initial state of the atmosphere, ocean, and biosphere to start the model calculations. The equations that describe the various interconnected processes in the earth system give rise to slightly different climate projections as the simulation progresses when slightly different initial conditions are used in the same model. An ensemble of model simulations with small changes in initial conditions could give rise to differences in projected surface temperature at a given location for up to 30 years from the start of the simulations (Deser et al. 2012). This internal variability tends to be higher as the geographical location over which this analysis is performed gets smaller. Internal variability is a dominant factor in the first few years after the start of the simulation, and decreases or becomes less important as the time of integration increases. Another way to understand this would be to think of this as day-to-day and year-to-year noise produced within the atmospheric system, which will be higher when we look at small spatial scales (e.g., a city) compared to a region (e.g., state scale).

Bias Correction

A critical step in using climate model projections for impacts assessments to correct for model bias. Bias corrections are usually applied when assessing hydrological impacts. The model bias is defined as the difference, when averaged over several years, between a chosen set of observations and calculated values from a model at the appropriate spatial scale. The choice of time period over which the average bias estimate is generated is constrained by the availability of observations, and a 30-year band covering the most recent historical period is used. Figure A1 is an example of this process. The left panel shows monthly average precipitation from observations over a model grid cell in Portland, Maine (orange, solid line); the green, dashed line represents the averaged model precipitation over the same region, averaged over the spring months, for 30 years. The figure on the right is another example of the model-observation difference, as a distribution of bias over the same 30-year period in the Great Plains (GP) region (in the form of a box plot). Each of the colored boxes represents a different model simulation for the 30 years, separated into the four seasons.

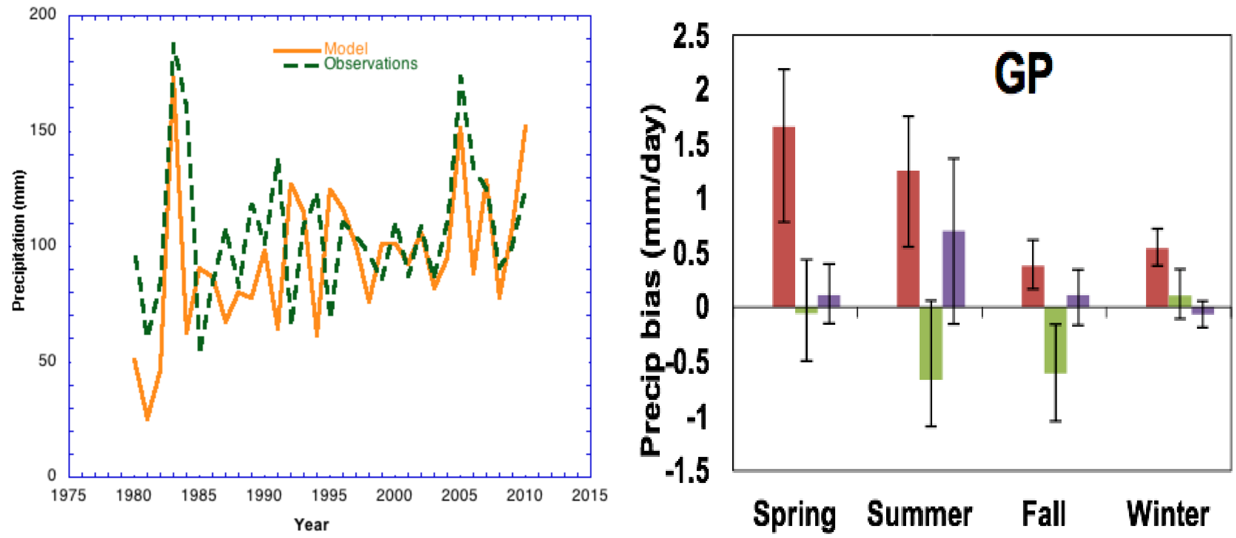

Figure A1. Process for Estimating Model Bias (left—average precipitation in spring over a single model grid cell [Portland, Maine] in mm from observations and model simulations; right—bias estimated from three model simulations over the four seasons as a box plot for the GP region; error bars denote the annual distribution of bias at the 10th and 90th percentiles).

Downscaling

All climate projections originate with coupled atmosphere-ocean GCMs. These models are driven by scenarios of future concentrations of GHG and other radiatively active substances; they generate projected changes in atmospheric, oceanic, and surface climate variables at scales that typically range from 100 to 300 km. Because these spatial scales are typically insufficient to accurately simulate regional conditions, generating climate projections at a regional level requires some method of downscaling. This downscaling is generally based either on an RCM (dynamic downscaling) or on empirical methods based on empirical approaches that use climate model outputs and climate observations (statistical downscaling).

Statistical Downscaling

Statistical downscaling can be relatively inexpensive, compared to the use of RCMs, when it is applied to just a few locations or with simple techniques. Statistical downscaling generally does not add any new information compared to the host climate model, and is suitable for generating quick assessments. It represents the process of obtaining fine grid output from coarse grid output by using statistical fits for current climate observations. This method can be tuned to obtain finer resolution output for targeted variables and for selected locations. This method's ease of use, and its flexibility, has led to a wide variety of applications for assessing impacts of climate change (e.g., Kattenberg et al. 1996; Hewitson and Crane 1996; Giorgi et al. 2001; Wilby et al. 2004, and references therein). Approaches encompass a range of statistical techniques, from simple linear regression (e.g., Wilby et al. 2000) to more complex applications based on weather generators (Wilks and Wilby 1999), canonical correlation analysis (e.g., von Storch et al. 1993), or neural networks (e.g., Crane and Hewitson 1998). These methods have successfully generated regional climate assessments for various governmental agencies and national reports. Figure A2 shows statistically downscaled results for the upper Mississippi catchment region produced using the Community Climate System Model (CCSM3.0) used in the Coupled Model Intercomparison Project (CMIP3) assessment. Two periods, 2021–2030 and 2051–2060, were used to perform statistical downscaling using approximately 130 weather stations in this region. Figure A2 also presents percent changes in precipitation for these two decades, compared to historical averages from observations.

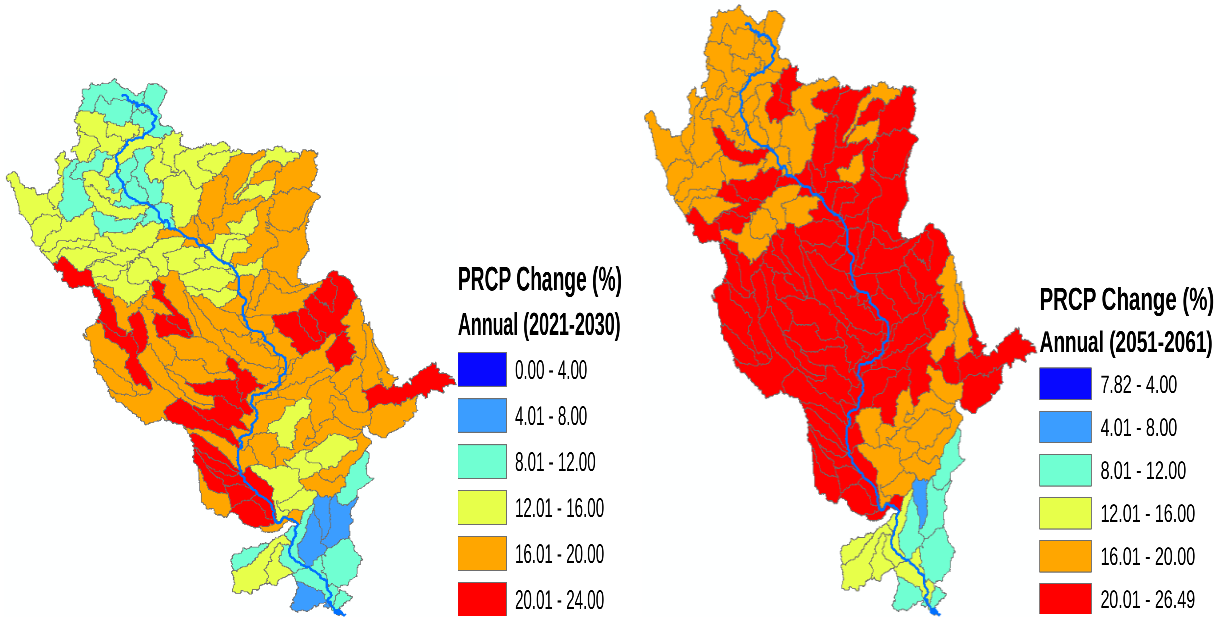

Figure A2. Statistical Downscaling of Precipitation for the Upper Mississippi Basin from a GCM (Franklin et al. 2010) (host climate output is from CCSM3.0 simulations for the A1B scenario generated for the IPCC AR-4 and archived on the Earth System Grid [ESG] website).

Dynamical Downscaling

Dynamical downscaling by RCMs generally refers to the use of limited area climate models that are forced at the boundaries using results from a host GCM (Giorgi and Mearns 1991, 1999; Wang et al. 2004; Liang 2005). These models were primarily developed by adapting mesoscale meteorological forecasting models to climate simulations. As a result, they have a full description of the land surface process, detailed cloud physics, and radiative transfer schemes (Giorgi et al. 2012). The higher spatial resolution of RCMs, as compared to GCMs, generally improves the ability to simulate climate, especially for fields such as precipitation that have high spatial variability. For example, some studies show that the higher RCM resolution yields better monsoon precipitation forecasts and interannual variability (Mo et al. 2005) and precipitation intensity (Roads et al. 2003). Thus, RCMs can effectively produce a more accurate forecast at regional scales in many instances. These models have been used widely in applications that require regional resolutions, and in particular where there is a need for higher-resolution climate projection for estimating hydrological vulnerabilities (Kenton et al. 2012; Chan et al. 2014; Mejia et al. 2012; Mearns et al. 2015).

Figure A3 shows the process for creating a dynamic downscaled product. The process starts by using the chosen RCM to perform a simulation over a time slice that has sufficient observational data to estimate bias in model calculations. The model requires three-dimensional preconditions for initialization and boundary conditions that will be regularly updated during the model simulation. Because the model only covers a fraction of the globe, the boundary conditions provide the inflow from regions outside the model domain into the model. These inputs are regularly updated so that the model experiences the same large-scale phenomena observed.

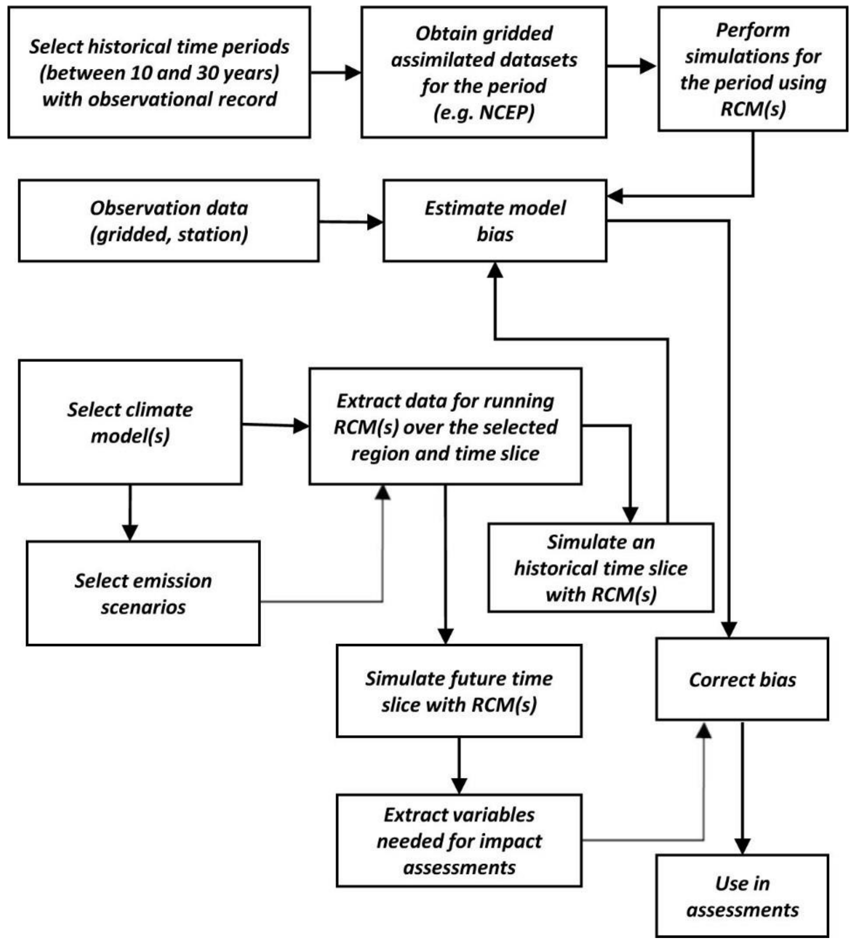

Figure A3. Dynamic Downscaling Process.

These historical model simulations are used to estimate the model bias. Observational datasets include observations from weather stations maintained by the National Weather Service. They are further processed to generate daily, monthly, and seasonal average observation files. The Precipitation-Elevation Regressions on Independent Slopes Model (PRISM) data is one such example. The next step involves selecting a future time slice over which to employ the RCM. Typically, RCMs are not used to simulate from present to 2100 as one continuous simulation; at present, this tends be computationally unfeasible at spatial resolutions lower than 50 km. Therefore, time slices of 10 years or more distributed around the mid-21st century and the end of the 21st century are typically chosen to generate the projection simulations. The input fields for these model runs are obtained from pre-existing global-scale simulations that are archived in repositories, such as CMIP5. These input conditions can be adjusted for estimated bias, as described earlier, before being input to an RCM or after the simulation results are obtained. One critical consideration for generating the downscaled model results is the choice of the GHG emission scenario for which the GCM results are available. The choice is made to meet the needs of the analysis and to match the number of such simulations that can be performed with available resources.

Future Climate Scenarios

To simulate future periods, the future conditions that affect weather and climate must be estimated and used as inputs in climate models. One of the conditions that is observed to be changing (increasing) is the concentration of GHGs in the atmosphere, which trap heat in a process similar to the one that causes high temperatures in a parked automobile on a sunny day. Car windows are transparent to the shorter wavelengths comprising visible sunlight, but opaque to the longer wavelengths of heat coming from surfaces heated by the sunlight. In the same way, gases like carbon dioxide and methane are transparent to visible light, but they absorb heat reflecting from the earth's surface. Both natural processes and human activity produce GHGs.

Climate scenarios are the set of conditions used as inputs to climate models to represent estimates of future conditions of GHGs. The files provided with this project include results from two selected future scenarios for two 10-year periods, and a historical 10-year period for comparison:

- RCP4.5: Representative Concentration Pathway 4.5 (human GHG emissions peak around 2040, then decline), with results for 2045 to 2054, and 2085 to 2094. RCP4.5 assumes that radiative forcing will increase by 4.5 W/m^2 by the year 2100, relative to pre-industrial values.
- RCP8.5: Representative Concentration Pathway 8.5 (emissions continue to rise throughout the 21st century), with results for 2045 to 2054, and 2085 to 2094. RCP8.5 assumes that radiative forcing will increase by 8.5 W/m^2 by the year 2100, relative to pre-industrial values.
- Historical: Based on historical conditions, with results for 1995 to 2004.

Boundary Conditions for Dynamic Downscaling

The regional-scale climate models used for dynamic downscaling cover only a portion of the earth. This makes it necessary to prescribe the inflow conditions into the model domain from regions outside the region covered by the regional climate model. These data usually come from either reanalysis data (assimilation of observations) or GCMs. This study uses three GCMs and one reanalysis dataset. The GCMs include Community Climate System Model Version 4 (CCSM4), which was developed by the National Center for Atmospheric Research, United States (Gent et al. 2011); GFDL-ESG2G, which was developed by the National Oceanographic and Atmospheric Administration (NOAA)/Geophysical Fluid Dynamics Laboratory, United States (Donner et al. 2011); and HadGEM2-ES, which was developed by Met Office Hadley Centre, United Kingdom (Jones et al. 2011). A combination of the GCM boundary conditions and the regional-scale model used leads to an individual ensemble member listed in Table 2. We use three GCMs to represent the range of response of GCMs to a doubling of atmospheric CO_2 from a baseline set in 1850. This response factor is known as the climate sensitivity of the model and is a critical parameter that distinguishes the more than 40 climate models used by the international community for climate change assessments. We use GFDL-ESM2G, which yields a global mean temperature change of 2.38°C with a doubling of CO_2; HadGEM2-ES, which has the highest sensitivity to a doubling of CO_2 with a projected temperature increase of 4.5°C; and CCSM4, which responds near the average of more than 30 GCMs with a temperature increase of about 2.9°C. These three GCMs provide the boundary conditions for our regional-scale simulations.

Additional ensemble members listed in Table 2 test the regional models' sensitivity to different approaches to running the model over the decadal time scales and any adjustments that were made to the boundary conditions obtained from GCMs, as explained earlier. To adjust the raw boundary GCM-provided conditions to generated "bias-corrected" simulations, we use the estimate of bias generated for a particular GCM with the regional-scale model compared to historical data. The mean values are bias adjusted, which preserves the variability present in raw climate model output in order to more closely match patterns in weather observation data. Additional details about this process appear in Wang and Kotamarthi (2015).

Nudging uses a correction term that depends on the difference between the model values and the boundary conditions. This correction term controls the strength of the nudging. Very strong nudging may destroy mesoscale patterns generated by the fine-scale models. Spectral nudging (SN) was applied in the zonal and meridional directions. Only the waves under a selected wavenumber, chosen to represent large-scale forcing, are retained in the nudging term. Here the waves being nudged are 1,000 km wide. Additional details about the nudging process appear in Wang and Kotamarthi (2013).

Description of Models

This study uses WRF version 3.3.1 to dynamically downscale three GCMs in one historical period (1995–2004) and two future periods (2045–2054 and 2085–2094) under RCP 4.5 and RCP 8.5. We apply the WRF model at a horizontal resolution of 12 km, with 600 west-east × 515 south-north grid points and 28 vertical levels over most of North America. The physics schemes used include the Grell-Devenyi convective parameterization (Grell and Devenyi 2002), Yonsei University planetary boundary layer scheme (Noh et al. 2003), Noah land surface model (Chen and Dudhia 2001), longwave and shortwave radiative schemes of the Rapid Radiation Transfer Model for GCM applications (http://rtweb.aer.com) (Lacono et al. 2008), and Morrison microphysics scheme (Morrison et al. 2009). Weak interior nudging is applied above 850 hPa to wavelengths around 1200 km, with a nudging coefficient of $3 \times 10^{-5}\ s^{-1}$. In both historical and future simulations, we allow a 1-year spin-up period for the model to reach equilibrium. Wang and Kotamarthi (2014) applied and tested these physics schemes and model setup for WRF driven by NCEP-R2 over the same model domain. Among these physics schemes, Morrison reduced the wet bias in winter over the GP and northern Rocky Mountains (see Figures 8a and 8c in Wang and Kotamarthi [2014]). Weaker nudging and a longer spin-up time reduced the wet bias in summer over the GP region. In addition, combining the weaker nudging with bias correction improved the RCM performance compared to only applying bias correction for the GCM (Xu and Yang 2015).

To assess the impacts of bias correction for the CCSM4 boundary conditions, we carried out two sets of dynamical downscaling: uncorrected CCSM4-driven WRF for the historical period and bias-corrected CCSM4-driven WRF for the late 21st century (Wang and Kotamarthi 2014). To explore the impacts of spectral nudging on model performance when bias correction is applied, we conducted two WRF runs driven by GFDL-ESG2G, with spectral nudging turned on in one of the simulations and turned off in the other (Zobel et al. 2017). Each set of simulations uses the same model setup and physics schemes, and they allow the same spin-up time.

Global Climate Models used for Boundary Conditions in This Study

The following is a description of the GCMs that provided the boundary conditions for the dynamic downscaling simulations listed in Table 1.

Community Climate System Model Version 4

The Community Climate System Model Version 4 (CCSM4) is a coupled global climate model, which includes atmospheric, land surface, ocean, and sea ice submodels that run simultaneously with a central coupler component. The University Corporation for Atmospheric Research (UCAR) developed CCSM4 with funding from the National Science Foundation, the Department of Energy, and the National Aeronautics and Space Administration. Results in this study from this model have raw boundary conditions, spectral nudging with a nudging strength of $3 \times 10^{-5}\ s^{-1}$, and a 1-year spin-up time.

This model is specified as `-model ccsm` in GATOR. Further details about the work done to generate these files are available in Wang and Kotamarthi (2015).

Community Climate System Model Version 4 (Bias Corrected)

A second set of results from the CCSM4 model are included in the input NetCDF files and computed results. Results in this study from this model have bias correction, spectral nudging with a nudging strength of 3 × 10-5 s-1, and a 1-year spin-up time (Wang and Kotamarthi 2015).

Use `-model ccsmbc` in GATOR to process the results of this model.

Geophysical Fluid Dynamics Laboratory Earth System Model Version 2G

The National Oceanic and Atmospheric Administration, Geophysical Fluid Dynamics Laboratory developed the Earth System Model Version 2G (ESM2G) (GFDL 2017). It includes an atmospheric circulation model and an oceanic circulation model, and takes into account land, sea ice, and iceberg dynamics. Processes simulated in the models include atmospheric chemistry, cloud physics, precipitation, evaporation, runoff, terrestrial ecology including carbon reservoirs, wave action, currents, sea ice, movement of fresh water though icebergs, ocean mixing, and marine biochemistry and ecology.

Results in this study from this model have bias correction, no spectral nudging, and a 1-year spin-up time.

Use `-model gfdlnn` in GATOR to process the results of this model.

Geophysical Fluid Dynamics Laboratory Earth System Model Version 2G (With Nudging)

A second set of results from the GFDL ESM2G bias-corrected model are included in the input NetCDF files and the computed results, with spectral nudging applied and a nudging factor of $3 \times 10^{-5} s^{-1}$, as documented in Zobel et al. (2017).

Use `-model gfdl` in GATOR to process the results of this model.

Hadley Global Environment Model 2 - Earth System

The United Kingdom's Met Office developed the Hadley Global Environment Model 2—Earth System (HadGEM-ES). It is used for both operational weather forecasting and climate research, and includes coupled atmosphere-ocean analysis and an earth system component that includes dynamic vegetation, ocean biology, and atmospheric chemistry. The atmospheric component includes either 38 or 60 vertical levels extending to approximately 40 km above the earth's surface, allowing it to simulate stratospheric processes and their influence on global climate. Earth system components simulate terrestrial and ocean carbon cycles, and tropospheric chemistry. Land vegetation and carbon are represented by five vegetation types (broadleaf tree, coniferous tree, grass adapted to cool season wet or dry conditions, grass adapted to warm/hot moist/dry conditions, and shrub). Modelled processes for ocean biology include effects of nutrients on plankton growth, and emissions of dimethylsulfide (DMS) to the atmosphere. DMS is believed to be a large source of sulfur going into the atmosphere, which increases cloud formation. Cloud cover increases cooling and would help offset the warming effects of GHGs.

HADGEM-ES will be used in the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. It is also one of the models used in the European ENSEMBLES project, which helps coordinate, validate, quantify, and reduce uncertainty, and maximize exploitation of the results of regional earth system models developed in Europe.

Results in this study from this model have raw boundary conditions, no spectral nudging, and a 1-year spin-up time.

Use `-model hadgem` in GATOR to process the results of this model.

National Centers for Environmental Predictions—Department of Energy, Reanalysis 2

The National Centers for Environmental Predictions—Department of Energy, Reanalysis 2 (NCEP-R2) is a joint data product from the National Centers for Environmental Prediction (NCEP) and the National Center for Atmospheric Research (NCAR), supported by the U.S. National Weather Service and the U.S. Department of Energy, Program for Climate Model Diagnosis and Intercomparison. It is a continually updated (1948–present), globally gridded dataset that represents the state of the earth's atmosphere, incorporating observations and numerical weather prediction model output from 1948 to the present. The reanalysis includes fixing errors and updating parameterizations of physical processes in the model.

Results in this study from this model include spectral nudging with a nudging strength of 3×10^{-4} s^{-1}, and a 1-day spin-up time. The project does not include modeling future time periods, so only results from the historical period are available.

Use `-model ncep` in GATOR to process the results of this data product.

Observed Meteorological Data

Daily precipitation, maximum temperature, and minimum temperature data are a gridded dataset based on observations (Maurer et al. 2002). This gridded dataset has a spatial resolution of 1/8 degree and covers 66 years, from 1950 to 2015. It has been applied extensively as a meteorological reference for evaluating dynamical and/or statistical downscaled results (e.g., Wood et al. 2004; Christensen et al. 2004; Maurer and Hidalgo 2008; Wehner 2013). The gridded precipitation within the continental United States is from the National Oceanic and Atmospheric Administration Cooperative Observer (co-op) stations. The precipitation gauge data are first gridded to the 1/8-degree resolution using the synergraphic mapping system algorithm of Shepard (1984), as implemented by Widmann and Bretherton (2000). The gridded daily precipitation data are then scaled to match the long-term average of the parameter-elevation regressions on independent slopes model (PRISM) precipitation climatology (Daly et al. 1994, 1997), which is a comprehensive dataset that is statistically adjusted to capture local variations due to complex terrain. The minimum and maximum daily temperature data over the contiguous United States, also obtained from co-op stations, are gridded using the same algorithm as for precipitation, and are lapsed to the grid cell mean elevation. We regridded this dataset onto the regional model's grid system to match the climate model grid domains, and for use with GATOR. The extent of these data are the contiguous United States, and unfortunately, data are not available for Alaska.

In GATOR, use `–model observed –scenario hist`, and `-statvariable precipobs`, `tempmaxobs`, or `tempminobs`. Estimates of model bias were computed with these data, and the results are included in the geographic information system database.

References

ACME (2014) *Accelerated Climate Modeling for Energy: Project Strategy and Initial Implementation Plan*, Washington, D.C., Department of Energy.

Chan, S.C., E.J. Kendon, H.J. Fowler, S. Blenkinsop, N.M. Roberts, and C.A.T. Ferro (2014) The Value of High-resolution Met Office Regional Climate Models in the Simulation of Multihourly Precipitation Extremes. *Journal of Climate* 27:6155–6174.

Chen, F., and J. Dudhia (2001) Coupling an Advanced Land Surface–Hydrology Model with the Penn State–NCAR MM5 Modeling System. Part I: Model Implementation and Sensitivity, *Monthly Weather Review*, 129, 569–585.Computing.io (2016) *WRF: Latitudes, Ellipsoids, and Datums*. Available at: http://computing.io/wp/2016/01/wrf-latitudes-ellipsoids-and-datums. Accessed 6/20/2017.

Christensen, N.S., A.W. Wood, N. Voisin, D.P. Lettenmaier, and R.N. Palmer (2004) The Effects of Climate Change on the Hydrology and Water Resources of the Colorado River Basin. *Climate Change* 62:337–363.

Crane, R.G., and B.C. Hewitson (1998) Doubled CO_2 Precipitation Changes for the Susquehanna Basin: Down-scaling from the GENESIS General Circulation Model. *International Journal of Climatology* 18:65–76.

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, doi:10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2.

Daly, C., G.H. Taylor, and W.P. Gibson (1997) The PRISM Approach to Mapping Precipitation and Temperature. *Proceedings of the 10th American Meteorological Society Conference on Applied Climatology*, pp. 20–23.

Deser, C., R. Knutti, S. Solomon, and A.S. Phillips (2012) Communication of the Role of Natural Variability in Future North American Climate. *Nature Climate Change* 2:775–779.

Donner, L.J., B. Wyman, R.S. Hemler, L.W. Horowitz, Y. Ming, M. Zhao, J.-C. Golaz, P. Ginoux, S.-J. Lin, M.D. Schwarzkopf, J. Austin, G. Alaka, W.F. Cooke, T.L. Delworth, S. Freidenreich, C.T. Gordon, S.M. Griffies, I.M. Held, W.J. Hurlin, S.A. Klein, T.R. Knutson, A.R. Langenhorst, H.-C. Lee, Y. Lin, B.I. Magi, S. Malyshev, P.C.D. Milly, V. Naik, M.J. Nath, R. Pincus, J.J. Ploshay, V. Ramaswamy, C.J. Seman, E. Shevliakova, J.J. Sirutis, W.F. Stern, R.J. Stouffer, R.J. Wilson, M. Winton, A.T. Wittenberg, and F. Zeng (2011) The Dynamical Core, Physical Parameterizations, and Basic Simulation Characteristics of the Atmospheric Component AM3 of the GFDL Global Coupled Model CM3. *Journal of Climate* 24(13), doi:10.1175/2011JCLI3955.1.

Flato, G., J. Marotzke, B. Abiodun, P. Braconnot, S.C. Chou, W. Collins, P. Cox, F. Driouech, S. Emori, V. Eyring, C. Forest, P. Gleckler, E. Guilyardi, C. Jakob, V. Kattsov, C. Reason, and M. Rummukainen (2013) *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Franklin, M., E. Yan, and Y. Demissie (2010) *Accounting for Downscaling and Model Uncertainties in Examining the Impacts of Climate Change on Hydrological Systems*. American Geophysical Union Fall Meeting, San Francisco, CA. San Francisco: AGU.

Gent, P.R., G. Danabasoglu, L.J. Donner, M.M. Holland, E.C. Hunke, S.R. Jayne, D.M. Lawrence, R.B. Neale, P.J. Rasch, M. Vertenstein, P.H. Worley, Z.-L. Yang, and M. Zhang (2011) The Community Climate System Model Version 4. *Journal of Climate* 24:4973–4991.

GFDL (Geophysical Fluid Dynamics Laboratory) (2017) *Earth System Models*. Available at: https://www.gfdl.noaa.gov/earth-system-model. Accessed 6/21/2017.

Giorgi, F., B. Hewitson, J. Christensen, M. Hulme, H. Von Storch, P. Whetton, R. Jones, L. Mearns, and C. Fu (2001) Regional Climate Information: Evaluation and Projections (Chapter 10). In: *Climate Change 2001: The Scientific Basis, Contribution of Working Group I to the Third Assessment Report of the IPCC.* [J.T. Houghton, Y. Ding, D.J. Griggs, M. Noguer, P.J. van der Linden, X. Dai, K. Maskell, and C.A. Johnson (eds.)]. Cambridge University Press: Cambridge, pp. 739–768.

Giorgi, F., E. Caoppola, F. Solomon, L. Mariotti, M.B. Sylla, X. Bi, N. Elguindi, G.T. Diro, V. Nair, G. Guiliani, U.U. Turuncoglu, S. Xozzini, I. Güttler, T.A. O'Brein, A.B. Tawfik, A. Shalaby, A.S. Zakey, A.L. Steiner, F. Stordal, L.C. Sloan, and C. Brankovic (2012) RegCM4: Model Description and Preliminary Tests over Multiple CORDEX Domains. *Climate Research* 52:7–29.

Giorgi, F., and L.O. Mearns (1991) Approaches to the Simulation of Regional Climate Change: A Review. *Review of Geophysics* 29(2):191–216.

Giorgi, F., and L.O. Mearns (1999) Introduction to special section: Regional Climate Modeling Revisited. *Journal of Geophysical Research* 104, doi: 10.1029/98JD02072.

Grell, G.A., and D. Devenyi (2002) A Generalized Approach to Parameterizing Convection Combining Ensemble and Data Assimilation Techniques, *Geophysical Research Letters* 29(14):1693, doi:10.1029/2002GL015311.

Hawkins, E., and R. Sutton (2009) The Potential to Narrow Uncertainty in Regional Climate Predictions. *Bulletin of the American Meteorological Society* 90:1095–1107.

Hewitson, B.C., and R.G. Crane (1996) Climate downscaling: Techniques and application. *Climate Research* 7:85–95.

Jones C.D., J.K. Hughes, N. Bellouin, S.C. Hardiman, G.S. Jones, J. Knight, S. Liddicoat, F.M. O'Connor, R.J. Andres, C. Bell, K.-O. Boo, A. Bozzo, N. Butchart, P. Cadule, K.D. Corbin, M. Doutriaux-Boucher, P. Friedlingstein, J. Gornall, L. Gray, P.R. Halloran, G. Hurtt, W.J. Ingram, J.-F. Lamarque, R.M. Law, M. Meinshausen, S. Osprey, E.J. Palin, L. Parsons Chini, T. Raddatz, M.G. Sanderson, A.A. Sellar, A. Schurer, P. Valdes, N. Wood, S. Woodward, M. Yoshioka, and M. Zerroukat (2011) The HadGEM2-ES Implementation of CMIP5 Centennial Simulations. *Geoscientific Model Development* 4:543–570, doi:10.5194/gmd-4-543-2011.

Kattenberg, A., Giorgi, F., Grassl, H., Meehl, G.A., Mitchell, J.F.B., Stouffer, R.J., Tokioka, T., Weaver, A.J., and Wigley, T.M.L. (1996). Climate Models - Projections of Future Climate. In: *Climate Change 1995: The Science of Climate Change*, 285–357 (J.T. Houghton, L.G.M. Filho, B.A. Callander, N. Harris, A. Kattenberg, and K. Maskell [eds.]) Cambridge University Press, Cambridge, UK.

Kenton, E.J., N.M. Roberts, C.A. Senior, and M.J. Roberts (2012) Realism of Rainfall in a Very High Resolution Regional Climate Model. *Journal of Climate* 25:5791–5806.

Lacono, M.J., J.S. Delamere, E.J. Mlawer, M.W. Shephard, S.A. Clough, and W.D. Collins (2008) Radiative Forcing by Long-lived Greenhouse Gases: Calculations with the AER Radiative Transfer Models, *Journal of Geophysical Research* 113:D13103, doi:10.1029/2008JD009944.

Liang, X.-Z., H. Choi, K.E. Kunkel, Y. Dai, E. Joseph, J.X.L. Wang, and P. Kumar (2005) Surface Boundary Conditions for Mesoscale Regional Climate Models. *Earth Interactions*, 9.

Maurer, E.P., and H.G. Hidalgo (2008) Utility of Daily vs. Monthly Largescale Climate Data: An Intercomparison of Two Statistical Downscaling Methods. *Hydrology and Earth System Sciences* 12:551–563, doi:10.5194/hess-12-551-2008.

Mearns, L.O., D.P. Lettenmaier, and S. McGinnis (2015) Uses of Results of Regional Climate Model Experiments for Impacts and Adaptation Studies: The Example of NARCCAP. *Current Climate Change Reports* 1(1):1–9.

Maurer, E.P., A.W. Wood, J.C. Adam, D.P. Lettenmaier, and B. Nijssen (2002) A Long Term Hydrologically Based Data Set of Land Surface Fluxes and States for the Conterminous United States. *J. Climate* 15:32373251. Available at: http://journals.ametsoc.org/doi/full/10.1175/1520-0442%282002%29015%3C3237%3AALTHBD%3E2.0.CO%3B2. Accessed on 6/20/2017.

Mejia, J.F., J. Huntington, B. Hatchett, D. Koracin, and R.G. Niswonger (2012) Linking Global Climate Models to an Integrated Hydrologic Model: Using an Individual Station Downscaling Approach. *Journal of Contemporary Water Research and Education* 147:17–27.

Mo, K.C., J.-K. Schemm, H.M.H. Juang, and R.W. Higgins (2005) Impact of Model Resolution on the Prediction of Summer Precipitation over the United States and Mexico. *Journal of Climate* 18:3910–3927.

Morrison, H., G. Thompson, and V. Tatarskii (2009) Impact of Cloud Microphysics on the Development of Trailing Stratiform Precipitation in a Simulated Squall Line: Comparison of One- and Two-moment Schemes. *Monthly Weather Review* 137:991–1007, doi:10.1175/2008MWR2556.1.

Noh, Y., W.G. Cheon, S.Y. Hong, and S. Raasch (2003) Improvement of the K-profile Model for the Planetary Boundary Layer Based on Large Eddy Simulation Data. *Boundary-Layer Meteorology* 107:401–427.

Roads, J., S.-C. Chen, and M. Kanamitsu (2003) U.S. Regional Climate Simulations and Seasonal Forecasts. *Journal of Geophysical Research Atmospheres* 108(D16), doi:10.1029/2002JD002232.

Shepard, D.S. (1984) Computer Mapping: The SYMAP Interpolation Algorithm. In: *Spatial Statistics and Models* (G.L. Gaile and C.J. Willmott [eds.]), Dordrecht, pp. 133–145.

von Storch, H., E. Zorita, and U. Cubasch (1993) Downscaling of Global Climate Change Estimates to Regional Scales: An Application to Iberian Rainfall in Wintertime. *Journal of Climate* 6:1161–1171.

Wang, J., and V.R. Kotamarthi (2015) High-resolution dynamically downscaled projections of precipitation in the mid and late 21st century over North America. *Earth's Future* 3:268–288, doi:10.1002/2015EF000304.

Wang, J., and V.R. Kotamarthi (2014) Downscaling with a nested regional climate model in near-surface fields over the contiguous United States. *Journal of Geophysical Research, Atmosphere* 119:8778–8797, doi:10.1002/2014JD021696.

Wang, J., and V.R. Kotamarthi (2013) Assessment of Dynamical Downscaling in Near-Surface Fields with Different Spectral Nudging Approaches Using the Nested Regional Climate Model (NRCM). *Journal of Applied Meteorology and Climatology* 52:1576–1591.

Wang, Y., L.R. Leung, J.L. McGregor, D.-K. Lee, W.-C. Wang, Y. Ding, and F. Kimura (2004) Regional Climate Modeling: Progress, Challenges, and Prospects. *Journal of the Meteorological Society of Japan* 82:1599–1628.

Wehner, M.F. (2013) Very Extreme Seasonal Precipitation in the NARCCAP Ensemble: Model Performance and Projections. *Climate Dynamics* 40:59–80, doi:10.1007/s00382-012-1393-1.

Widmann, M., and C.S. Bretherton (2000) Validation of Mesoscale Precipitation in the NCEP Reanalysis Using a new Gridcell Dataset for the Northwestern United States. *Journal of Climate* 13:1936–1950, doi:10.1175/1520-0442(2000)013<1936:VOMPIT>2.0.CO;2.

Wilby, R.L., S.P. Charles, E. Zorita, B. Timbal, P. Whetton, and L.O. Mearns (2004) *Guidelines for Use of Climate Scenarios Developed from Statistical Downscaling Methods*. Intergovernmental Panel on Climate Change (IPCC) Task Group on Data and Scenario Support for Impacts and Climate Analysis (TGICA). Available from http://ipcc-ddc.cru.uea.ac.uk/guidelines/StatDown_Guide.pdf.

Wilby, R. L., L.E. Hay, W.J. Gutowski, R.W. Arritt, E.S. Takle, Z. Pan, G.H. Leavesley, and M.P. Clark (2000) Hydrological Responses to Dynamically and Statistically Downscaled Climate Model Output. *Geophysical Research Letters* 27(8):1199–1202.

Wilks, D.S., and R.L. Wilby (1999) The Weather Generation Game: A Review of Stochastic Weather Models. *Progress in Physical Geography* 23:329–357.

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. *Climate Change* 62:189–216, doi:10.1023/B:CLIM.0000013685.99609.9e.

Xu, Z., and Z.-L. Yang (2015) A New Dynamical Downscaling Approach with GCM Bias Corrections and Spectral Nudging. *Journal of Geophysical Research: Atmospheres* 120:3063–3084, doi:10.1002/2014JD022958.

Zobel, Z., J. Wang, D.J. Wuebbles, and V.R. Kotamarthi (2017) Evaluations of High-Resolution Dynamically Downscaled Ensembles over the Contiguous United States. *Climate Dynamics,* in press. doi:10.1007/s00382-017-3645-6.

Appendix B: GIS Database of Example Results

The following eight sets of statistics, described in the Results section of this report, are available with the project files as a set of ESRI File Geodatabases in the GATOR\Data\GIS\Geodatabases folder. An ESRI ArcGIS 10.3 ArcMap project is also provided with symbolized layers for each statistic. ESRI File Geodatabases are also readable through the application programming interface available at http://appsforms.esri.com/ products/download/#File_Geodatabase_API_1.3, and most open-source GIS tools support ESRI file geodatabase format. Table B1 summarizes the contents and naming conventions of the geodatabases and the feature datasets within them in the GIS folder.

Table B1. Geodatabase and Feature Dataset Naming Conventions for the Eight Statistics Calculated

	Geodatabase/Feature Dataset (FD) Name			
Statistic	**WRF Projection Data for Contiguous United States**	**WRF Projection Data for Alaska**	**Standard Projection Data for Contiguous United States[1]**	**Standard Projection Data for Alaska[2]**
Number of Days Over 90°F	WRF_Outputs.gdb/ TempF_DyGt_90	WRF_AK_Outputs.gdb/ TempF_DyGt_90	TempF_DyGt_90.gdb/ Contiguous_US	TempF_DyGt_90.gdb/Alaska
Number of Days with a Heat Index Over 90° F	WRF_Outputs.gdb/ HtIdx_DyGt_90	WRF_AK_Outputs.gdb/ HtIdx_DyGt_90	HtIdx_DyGt_90.gdb/ Contiguous_US	HtIdx_DyGt_90.gdb/Alaska
Heat Waves	WRF_Outputs.gdb/ TempF_GtWave90_3Day	WRF_AK_Outputs.gdb/ TempF_GtWave90_3Day	TempF_GtWave90_3Day.gdb/ Contiguous_US	TempF_GtWave90_3Day.gdb/ Alaska
Monthly and Annual Precipitation	WRF_Outputs.gdb/ PrecipIn_Total	WRF_AK_Outputs.gdb/ PrecipIn_Total	PrecipIn_Total.gdb/ Contiguous_US	PrecipIn_Total.gdb/Alaska
Drought	WRF_Outputs.gdb/ PrecipIn_ConsecDyLe_0	WRF_AK_Outputs.gdb/ PrecipIn_ConsecDyLe_0	PrecipIn_ConsecDyLe_0.gdb/ Contiguous_US	PrecipIn_ConsecDyLe_0.gdb/ Alaska
Extreme Precipitation	WRF_Outputs.gdb/ PrecipIn_Percentile_90	WRF_AK_Outputs.gdb / PrecipIn_Percentile_90	PrecipIn_Percentile_90.gdb/ Contiguous_US	PrecipIn_Percentile_90.gdb/ Alaska
Multi-Model Averages	WRF_Outputs.gdb/ <above FDs per statistic>	WRF_AK_Outputs.gdb/ <above FDs per statistic>	Model_Averages.gdb/ Contiguous_US	Model_Averages.gdb/Alaska
Model Bias	WRF_Outputs.gdb/ <above FDs per statistic>	WRF_AK_Outputs.gdb/ <above FDs per statistic>	Model_Bias.gdb/ Contiguous_US	N/A[3]

[1] USA contiguous Albers equal area conic U.S. Geological Survey.
[2] Alaska Albers equal area conic, 1983 North American Datum.
[3] Observed data are not available for Alaska.

Names of the GIS layers in the geodatabases, and field names within them, are intended to be self-documenting. They match the conventions used for the GATOR parameters and options. Some examples are provided below:

- The TempF_DyGt_90_ccsm_hist layer contains temperature statistics in Fahrenheit units, with the statistic being days with temperatures higher than 90°F. The files used for processing were from the ccsm set (raw CCSM4 boundary conditions) and the historic period of 1995 to 2004.
- Within the layer is a series of DyGt_TempF_90_YYYY fields containing per-year results (YYYY indicates each year), and a final DyGt_TempF_90_1995-2004 field with the cumulative 10-year results.
- The PrecipIn_Total_gfdlnn_rcp45mid layer contains precipitation statistics in units of inches, with the statistics representing total monthly and annual precipitation. The files used for processing were from the gfdlnn set (GFDL ESM2G boundary conditions with no nudging applied), the midcentury period of 2045 to 2054, and the RCP4.5 scenario.

- Within the layer is a series of TotalPrecipIn_MMMYYYY fields containing monthly year results for each year (MMM indicates each month and YYYY indicates each year), a TotalPrecipIn_YYYY field for each year containing the annual totals, a set of MeanTotal_PrecipIn_MMM2045_2054 fields containing the monthly totals averaged over 10 years, and finally a MeanTotal_PrecipIn_2045_2054 field containing the average of the 10 annual totals.
- The TempF_GtWave90_3Day_ModelAverage_rcp85end layer contains temperature statistics in degrees Fahrenheit; the statistic is the number of times the maximum daily temperature was greater than 90°F for at least 3 consecutive days. ModelAverage indicates that the results from all the available models were averaged, and the results are for end-of-century period of 2085 to 2094, and the RCP8.5 scenario.
- Within the layer is a series of "ccsm," "ccsmbc," and related fields containing the computed statistic for that model, and a final GtWave_TempF_90_3Day_ModelAverage_2085_2094 field containing the average of the available models.

If GATOR runs for less than the available 10 years, the names of the cumulative fields reflect the shorter range, and the averages take into account the number of years processed. If only 1 year is processed, the cumulative fields are omitted. All of the computed statistics used the full 10-year periods available, except for a few special cases where files are only available for 6 of the 10 years. For those exceptions, an estimate of the 10-year field was manually added by assuming the 10-year result is proportional to the available data.

The ESRI ArcGIS 10.3 ArcMap project file (GATOR\Data\GIS\Results.mxd) contains maps of all the computed layers, symbolized with consistent ranges and colors for each statistic type. The layers are organized into contiguous United States and Alaska sets because of the different projections used. They are grouped into sections named for each statistic type. If ArcGIS 10.3+ is installed, simply double-click the project file to open it.