**Towards a data hub of global daily weather observations for the ERS**

**Noé J Nava** ([Noe.Nava@usda.gov](mailto:Noe.Nava@usda.gov)), and Maros Ivanic ([Maros.Ivanic@usda.gov](mailto:Maros.Ivanic@usda.gov)): APM /MTED/ERS/USDA

**Collaborator: Administrative Data Research Facility (ADRF)**: Secured cloud-based computing platform designed to promote collaboration, facilitate documentation, and provide information about data use to the agencies that own the data.

Quick outliner about short technical document:

* Brief description of database and its intended use
* How NASA’s spatial/temporal data can be manipulated in R
* Description of NASA database and our needs

**Brief description of database and its intended use**

APM proposed to centralize weather data sources by creating a data hub for MTED economists to collect historical and forecasted weather data, facilitating analysis replication, and providing internal and external validity to ERS outputs, including reports and journal articles. A long-standing problem in the agricultural economics is the employment of data sources that vary in their methodologies and ease of access. The PRISM Climate Group at Oregon State has successfully standardized a process to provide historical data. Because much of the agricultural economist’ job relates to projections and forecasts that depend on future weather, there is need to have a similar standardization of forecasted weather data.

Recognizing the previous limitation, the Intergovernmental Panel on Climate Change (IPCC), collaborating with NASA Center for Climate Simulation, put together the NASA Earth Exchange Global Daily Downscaled Projections (NEX-GDDP-CMIP6). NEX-GDDP-CMIP6 forecasted scenarios follow the IPCC guidelines for a business as usual (1.5 – 3.0 ℃ average global temperature). Because weather projections depend on Shared Socioeconomic Pathways (SSPs) assumptions, are highly granular in their spatial resolution (0.25 degrees x 0.25 degrees), extend over large periods of time (2014 – 2100), and consider several weather variables, even a useful subset of NEX-GDDP-CMIP6 can be in the several TBs, and their manipulation not practical. Below, we propose a subset of the dataset that is both helpful and practical for ERS business.

**How NASA’s spatial/temporal data can be manipulated in R**

For a step-by-step demonstration of how to use NASA data, see the following repository:

<https://github.com/Noe-J-Nava/ADRF_example>

To follow that work, please download the following .nc file (about 250MB, exceeding by 150MB GitHub limit), and place it in the data directory:

<https://portal.nccs.nasa.gov/datashare/nexgddp_cmip6/CanESM5/ssp126/r1i1p1f1/tasmin/tasmin_day_CanESM5_ssp126_r1i1p1f1_gn_2015.nc>

**Description of NASA database and our needs**

NASA temporal data sharing portal: https://portal.nccs.nasa.gov/datashare/nexgddp\_cmip6/.

Database is comprised of yearly and daily netCDF geospatial data structures. NetCDF, indexed by .nc, are global projections that are gridded by longitude and latitude. All of the daily observations have the same dimensions. However, each of these NetCDF can only represent one single variable. Thus, projections represent the value of the variable on the plane.

A data transformation from global projections to, for instance, U.S. counties requires a second projection of the geographical region to be studied. A U.S. map is imposed over the global map, and an algorithm identifies the overlapping regions to calculate regional average. There exist several R and Python packages that do this task efficiently.

Below I further describe NASA database, and propose a subset of their data that fits ERS needs.

|  |  |
| --- | --- |
| **NEX-GDDP-CMIP6** | **ERS-Subset** |
| *Spatial coverage* | |
| West Bounding Coordinate: 180 W  East Bounding Coordinate: 180 E  North Bounding Coordinate: 90 N  South Bounding Coordinate: 60 S | West Bounding Coordinate: 180 W  East Bounding Coordinate: 180 E  North Bounding Coordinate: 90 N  South Bounding Coordinate: 60 S |
| *Temporal resolution and extent* | |
| Daily from 2014-01-01 to 2100-12-31 | Daily from 2021-01-01 to 2060-12-31 |
| *Weather variables* | |
| 1. Near-Surface Relative Humidity 2. Near-Surface Specific Humidity 3. Precipitation 4. Surface Downwelling Longwave Radiation 5. Surface Downwelling Shortwave Radiation 6. Daily-Mean Near-Surface Wind Speed 7. Daily Near-Surface Air Temperature 8. Daily Maximum Near-Surface Air temperature 9. Daily Minimum Near-Surface Air Temperature | 1. Precipitation 2. Daily Maximum Near-Surface Air temperature 3. Daily Minimum Near-Surface Air temperature |
| *Shared Socioeconomic Pathways* | |
| 1. SSP1-2.6: Sustainability (Low challenges to mitigation and adaptation) 2. SSP2-4.5: Middle of the Road (Medium challenges to mitigation and adaptation) 3. SSP3-7.0: Regional Rivalry (High challenges to mitigation and adaptation) 4. SSP5-8.5: Fossil-fueled Development (high challenges to mitigation, low challenges to adaptation) | 1. SSP2-4.5 2. SSP3-7.0 |
| *Climate laboratories* | |
| Not clear in NASA documentation, but it seems that they partnered with about 30, so each data laboratory is a dataset as previously described. | To minimize the problem of laboratory errors, we would need to average across 4-8 laboratories. Such averaging is standard practice in applied economics papers.  Notice here that, if possible, we do not need to have all laboratories but only one dataset representing all of them. |