The DASH Portal: Supporting Agricultural Research by Automating Geospatial Data Tasks

Heather M. Savoy1, Debra P.C. Peters1, Geoavny A. Ramirez2, and N. Dylan Burruss2

1 USDA ARS, Jornada Experimental Range Unit, and Jornada Basin Long Term Ecological Research Program, Las Cruces, NM, USA
2 New Mexico State University, Jornada Experimental Range Unit, and Jornada Basin Long Term Ecological Research Program, Las Cruces, NM, USA

Objective
Continental-scale geospatial data, such as daily precipitation or soil properties across North America, have the capability to serve the data needs of a wide variety of agro-ecological projects, but their cost in terms of time and memory can hinder efficient integration. To reduce the time of downloading big data sets from online repositories, cropping the data to desired extents, harmonizing the resolutions and coordinate reference systems, and performing statistical analyses for multiple data sets, we aim to create a portal that automates these steps.

DASH: Data Access and Spatiotemporal Harmonization

Motivation: Agroecosystem Studies

Ecology of Complex Disease Systems
- A USDA Grand Challenge project was created to use Vesicular Stomatitis Virus (VSV) as a model for predictive disease ecology.
- The disease ecology system for VSV in a complex network of interactions among the virus, multiple insect vectors, multiple livestock hosts, and the environment. A big data-model integration approach is being used for this transdisciplinary project (Peters et al., 2018, 2020).
- Predictive models are being built to serve as an early warning system (Peters et al., submitted), and these models incorporate many geospatial data layers, such as: (a) Climate (precipitation, temperature) (PRISM, NOAA), (b) Vegetation (LANDSAT), (c) Horse density (USDA), and (d) Distance to streams with water (USGS).
- The DASH portal will serve as a data repository for this project to help scientists to access both the original and processed geospatial data layers used in this project.

Ecology of the Dust Bowl
- This study is using historical agricultural data to predict ecological consequences of changes in multi-scale connectivity, climate, and land use (Peters et al., 2020).
- Data records from the 1933-1939 large-scale drought in the Central US, as well as before (1926-1932) and after (1940-1949), are being used to relate changes in agricultural production to climate, landscape, and broad-scale features of the land surface.
- Geospatial data involved include: (a) County-level corn yield and land use records, (b) Climate (precipitation, temperature), (c) Soil texture properties, (d) Simulated soil water content, (e) Broad-scale dust and sand movement, and (f) Landform distribution.
- The DASH portal will host the organized historical data and the derived variables from SOILWAT simulations for use by other scientists.

Comparisons
- **Geographic Information Systems (GIS)** software programs exist to perform many of the operations in this tool, but the over-abundance of features can slow down the programs. Processing large datasets in GIS can cause unreasonable processing times or software crashes. This portal strives to support scientists needing large datasets or who potentially do not have GIS experience.
- **Online Data Portals** exist to host different datasets. For example, The National Map (https://nationalmap.gov) provides multiple data layers for the US, but they are topography related and downloading multiple layers may result in different spatial extents. This portal will contain data from many fields to support complex, multi-faceted agro-ecosystem research and provide fully harmonized data cubes for downloading.
- **Google Earth Engine (GEE)** is a powerful tool for server-side geospatial processing and large dataset access. However, performing those processes requires knowledge of a GEE variant of JavaScript or Python. Download options are limited to Google Drive or Google Cloud, which could incur fees for large datasets. This portal does not require any programming skill or fees.

References