TEXAS A&M Institute of Data Science

Texas A&M Institute of Data Science Seminar Series

Large-scale Geospatial Data Analysis and Visualization in R

Summary

Analysis of large-scale geospatial data (regional- to global scale at high spatial and temporal resolution) can be computationally expensive and time-consuming, especially when working with multiple formats and sources of data. R provides a powerful computational alternative to popular Geographic Information System (GIS) software to organize, analyze and visualize geospatial data. R enjoys a vast collection of open-source libraries for GIS-type operations and proven statistical analysis and data visualization capabilities. Taking examples from global satellite data in gridded/ raster format, we will demonstrate several geospatial operations like projections, resampling, spatial extraction, cropping, masking etc. using rasters, shapefiles, and spatial data frames. For a seamless analysis across different data types and platforms, conversion from/to different data formats like data frames, matrices, rasters, and structured data like NetCDF will be discussed. Advanced topics will include working with data cubes (terra spatrasters), layer-wise operations on data cubes, cell-wise operations on raster time series by implementing user-defined functions with *stackApply*. Multicore parallel application (*parApply* and *clusterR* family) of user-defined functions in R will be demonstrated for large-scale dataset.

- **Instruction type**: Hands-on.
- **Course level**: Intermediate
- **Course requirements**: Computer with installed and working R and RStudio (*a.k.a.* Posit, https://posit.co/download/rstudio-desktop/). Working experience in R is required.

- **Format**: Hybrid (In-person + Online)

Learning Objectives

- Plotting and visualizing geospatial data
- Efficient GIS and statistical operations on large geospatial dataset
- Application of custom functions for large-scale gridded/ spatial dataset analysis
- Parallelization of geospatial operations in R

Time: 10:15 a.m.-12:45 p.m. US Central Time (Wednesday, Oct 25th, 2023)
In-person Location: BLOC 220
Zoom Link: https://tamu.zoom.us/j/94200148283?pwd=aCs5SjdZbFAzb2xnekpFQ1VmNmFWdz09

Host: Debasish Mishra, Data Science Ambassador@TAMIDS, debmishra@tamu.edu
Leah Kocian, PhD Student@BAEN, leahkay4@tamu.edu

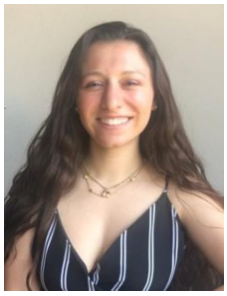
HIGH PERFORMANCE RESEARCH COMPUTING TEXAS A&M UNIVERSITY
TEXAS A&M Institute of Data Science
TEXAS A&M UNIVERSITY Biological & Agricultural Engineering

TEXAS A&M Institute of Data Science

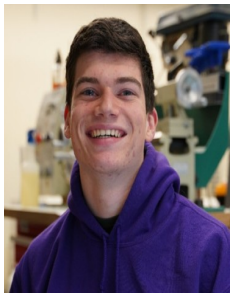
Texas A&M Institute of Data Science Seminar Series

Instructors

Debasish Mishra is a Ph.D. student at the Biological and Agricultural Engineering Department at Texas A&M University. Debashish studies the coupling of global scale water and energy cycles, and inferring predictive skills for flash floods and droughts. His research interests include hydro-climatology, soil science, remote sensing, data fusion, and Machine Learning.



Leah Kocian is a Ph.D. student at the Biological and Agricultural Engineering Department at Texas A&M University. Her research is centered on the investigation of contaminant fate and transport within urban garden root zones, with a specific focus on its potential implications for human health. Her methodology encompasses a suite of experimental techniques. She supplements these methods with advanced modeling and machine learning approaches by utilizing existing functions and developing new ones in R.



Aidan Brown is a 4th year undergraduate at Texas A&M University majoring in Geographic Information Science and Technology. He has research experience in geospatial analysis, remote sensing, and statistical modeling. He is interested in large-scale data analysis, geospatial image processing, and environmental modeling.