Outlook

Meeting assets for GeoComp & ML 2025 course are ready!

From Zoom <no-reply@zoom.us>
Date Tue 9/30/2025 10:11 AM
To Amatulli, Giuseppe <giuseppe.amatulli@yale.edu>

zoom

Meeting assets for GeoComp & ML 2025 course are ready!

Meeting summary

Quick recap

The meeting began with troubleshooting software installation issues for GW384 programs, which were successfully resolved through technical guidance. The session then focused on data processing techniques for large datasets, particularly for the United States portion of GeoCom analysis, where Giuseppe demonstrated various methods for efficient data handling and file management. The final portion covered advanced geospatial data processing techniques using GDAL tools, including methods for handling raster files, working with Virtual Raster Tiles, and efficient data extraction from large text files, with discussions about the balance between using base languages versus higher-level programming environments.

Next steps

- All attendees to practice using GDAL commands for their geospatial data processing tasks.
- All attendees to review the GSIM assignment solution.
- All attendees to watch the video recording on multi-core operations with GDAL and PK Tools.
- All attendees to start organizing their geodata for their final work using GDAL commands.
- All attendees to prepare their datasets for machine learning procedures.
- Olha to try using the correct syntax for GW384 version programs like GeoEdit and GDELCalc.
- All attendees to explore using VRT for efficient file management.
- All attendees to practice using GDAL location info for extracting information at specific coordinates.

Summary

GW384 Software Troubleshooting and Analysis

The meeting focused on troubleshooting issues with GW384 software versions, particularly for programs like Geo Edit and GDELCalc, which Olha reported were not functioning properly. Giuseppe

guided Olha through checking installation and file path issues, and the problem was resolved. The session then shifted to covering the second part of the GeoCom analysis, where Giuseppe explained a solution involving efficient data processing techniques for large datasets, particularly focusing on the United States portion of the data. He demonstrated methods for creating date files, handling IDs, and using join commands while emphasizing the importance of proper file sorting for successful data merging.

Data Processing Solutions with AWK

Giuseppe explained different solutions for processing data files, focusing on handling NA values and efficient file creation. He demonstrated how to use AWK to process data without including NA values and showed a more efficient solution using associative arrays to create multiple files simultaneously based on dates. Giuseppe also addressed questions from Autumn and Marianne about the use of dollar signs to reference columns and the difference between appending to a text file versus writing to a text file in different solutions.

GDAL Tools for Geospatial Data

The meeting focused on using GDAL tools for geospatial data processing, with Giuseppe explaining various techniques to manipulate raster files. He demonstrated how to work with AWK for efficient data processing and emphasized the importance of aligning pixel values and data types to avoid shifts in datasets. Giuseppe also covered how to use GDAL to change projections, handle no-data values, and work with large datasets by breaking them into tiles. The session concluded with instructions on using GDAL to extract pixel values at specific locations, including methods for handling multiple points from text files.

Efficient Latitude-Longitude Data Extraction

Giuseppe demonstrated a method for efficiently extracting latitude and longitude data from large text files using the paste command in a loop, which is faster than using Rasterio or other methods for large datasets. He explained that this approach avoids loading the entire file into memory and instead processes data directly from the file, though he noted that care must be taken when dealing with points outside the data processing area, which can result in empty lines. Marianne asked about the need for the left-pointing arrow in front of statements, and Giuseppe clarified that it is necessary for the paste command to work correctly when processing multiple files.

Understanding Zero-Indexed Paste Function

Giuseppe explained how to use the paste function in a coding context, emphasizing the importance of specifying the correct format to avoid errors. He demonstrated how to handle file locations and image coordinates, clarifying that these values are zero-indexed. Quinn asked for clarification on certain code elements, and Giuseppe provided explanations about matrix rows and columns, as well as pixel coordinates.

TIFF to Text Conversion Techniques

Giuseppe explained how to convert TIFF files to text format using GDAL Translate for error checking and manipulation, demonstrating how to use AWK for thresholding pixel values. He also introduced Virtual Raster Files (VRTs) as a way to efficiently handle large datasets by creating pointers rather than duplicating data, allowing for faster operations and better disk space management. Giuseppe emphasized the importance of mastering VRTs for efficient multi-core processing and handling

overlapping images, though he noted that proper use requires careful consideration of options and practice.

VRTs in GIS Analysis

Giuseppe explained the use of Virtual Raster Tiles (VRT) in GIS analysis, highlighting their efficiency in handling large datasets and their compatibility with various software like R and Python. He demonstrated how VRTs can be used for multi-core processing and emphasized their utility in high-performance computing environments. Giuseppe also introduced the concept of using shapefiles to manage spatial data efficiently and discussed the OGR command for vector file operations. Quinn asked about the balance between using base languages like GDAL and higher-level languages like Python or R, to which Giuseppe replied that he prefers conducting GIS and remote sensing analyses in Bash using GDAL and PK tools, reserving Python for modeling tasks.

Thank you,

Zoom Support Team
https://support.zoom.us

Zoom.com
55 Almaden Blvd
San Jose, CA 95113

+1.888.799.9666