O Outlook

Meeting assets for GeoComp & ML 2025 course are ready!

From Zoom <no-reply@zoom.us>
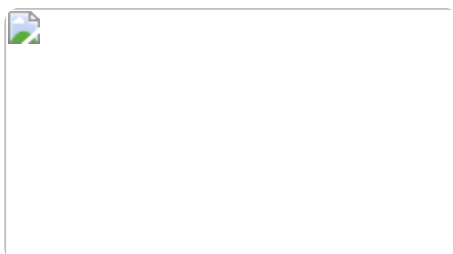**Date** Thu 9/25/2025 10:09 AM
To   Amatulli, Giuseppe <giuseppe.amatulli@yale.edu>

ZOOM

Meeting assets for GeoComp & ML 2025 course are ready!

Recording

Duration: 02:07:29

Share link: https://yale.zoom.us/rec/share/v2fNDtJUaLbF
mRrQKJcZ7ZCVeW2fbKTMjM_y44R7c_fi2oWAAgXITWX
pddCmGqvg.Ai3CT9W0ki8zXoGj

[Watch recording]   [Share]

Cloud recordings will be deleted automatically after they have been stored for 395 days.

Meeting summary

Quick recap

The meeting focused on geographic data processing and manipulation techniques, with extensive discussion of raster data handling using tools like GDAL and OpenEV. Giuseppe provided detailed instruction on various data processing concepts including file formats, data types, and handling no-data values, while also covering practical applications and best practices for working with geographic data. The session concluded with guidance on file compression, resampling techniques, and alternative data storage formats, along with recommendations for using existing published datasets rather than building custom mosaics.

Next steps

- Quinn: Discuss with boss about using LiDAR data as a spatial benchmark for their forest modeling work, particularly to estimate post-fire stand densities and tree heights.

- Quinn: Develop a project proposal for the Akaton week in Matera, focusing on forest structure analysis using spatial data.
- Quinn: Explore OpenTopography.org for LiDAR data in Alaska to potentially estimate post-fire stand densities and tree heights.
- Giuseppe: Discuss GSIM dataset assignment solutions in next Tuesday's session.
- Giuseppe: Follow up with students who signed up for the GSIM dataset assignment for discussion next Tuesday.
- Quinn: Develop a clear project goal to justify remote work.
- Giuseppe: Provide more information on accessing and extracting data from NetCDF subdatasets.
- Quinn: Consider visiting Matera for the project work.

Summary

LiDAR Project Planning Discussion

Quinn and Giuseppe discussed Quinn's potential participation in an upcoming Akaton week in Matera, where Quinn expressed interest in developing a project involving forest structure analysis in Alaska using LiDAR data from OpenTopography.org. They explored the possibility of estimating sand densities and tree heights using LiDAR, which Giuseppe confirmed could be feasible. Quinn also shared her initial attempts at manipulating GSIM dataset files, which will be discussed further in next week's session.

Efficient Data Processing Techniques

The meeting focused on the use of Bash and AWK for file management and text manipulation, highlighting their power for multi-core computations and data manipulation. Giuseppe discussed the importance of standardizing raster data for efficient processing in Python, R, and GRASS, emphasizing the role of GDAL and OGR in handling geographical data. He introduced the concept of hierarchical data formats like NetCDF and HDF, and explained the use of CDO Climate for efficient NetCDF processing. Giuseppe also covered GDAL commands, including gdalinfo, gdaltranslate, and gdalwarp, and explained the significance of projection codes and metadata in geographical data processing.

Raster Data Manipulation Techniques

Giuseppe led a discussion on raster data manipulation and alignment, emphasizing the importance of matching pixel resolution and origin points to avoid errors. He introduced various GDAL tools for tasks such as metadata editing, raster warping, and point extraction, highlighting JIRA Location Info as the fastest method for large datasets. Mikulas raised a question about efficiently assigning raster data to points, and Giuseppe recommended using JIRA Location Info for faster processing. Sofia inquired about cropping rasters, and Giuseppe advised being cautious with software behavior and suggested transforming vectors to rasters before cropping. The session concluded with a brief overview of GDAL tutorials and exercises, encouraging participants to explore the tools and their Python API for practical applications.

OpenEV vs QGIS TIFF Viewing

Giuseppe discussed the advantages of using OpenEV over QGIS for viewing TIFF files, highlighting its speed and ability to display pixel values accurately without metadata interference. He explained the importance of understanding the pixel size, projection, and no-data values, emphasizing that incorrect handling of these details can lead to computational errors, especially in high-resolution

data. Giuseppe also provided tips on checking and correcting pixel resolution approximations to ensure accurate alignment and computation of TIFF files.

Data Types in Geographic Systems

Giuseppe explained the importance of data types in computer systems, particularly when working with geographic data. He discussed how different data types have specific ranges and limitations, such as the byte type which can only store values from 0 to 255. Giuseppe also covered how to handle situations where the data exceeds the range of a particular data type, such as using unsigned 16-bit integers for values between 0 and 65535. He emphasized the importance of choosing the appropriate data type to balance precision and file size, and provided examples of how to scale and offset values to fit within desired ranges.

Manipulating No Data Values in Raster Data

Giuseppe explained the concept of no data values in raster data, emphasizing that they are typically assigned an extreme value and labeled as "no data" to indicate missing or invalid data. He demonstrated how to use GDAL Edit to change the no data value label while keeping the pixel value unchanged, and discussed the importance of correctly setting no data values for proper visualization and statistical analysis. Giuseppe also addressed questions from Quinn and Angelo about the use of special characters for no data in programming languages and the effects of changing pixel resolution in GDAL Edit.

Understanding New File Concepts

Giuseppe and Hager discussed file types and policies that were new to Hager. Giuseppe suggested that the best way to understand these concepts was to try them out. Hager expressed uncertainty about the vasters and papers, prompting Giuseppe to acknowledge the knowledge gap and offer support.

Raster vs Vector GIS Data

Giuseppe explained the differences between raster and vector data types in geographic information systems, emphasizing the use of matrices for rasters and polygons/lines/points for vectors. He discussed the importance of correctly identifying and saving data types when working with GDAL tools and other software, providing examples and tips for handling data types in calculations and output. Giuseppe also highlighted the significance of aligning datasets for accurate matrix operations and the use of tools like AWK for extracting specific data.

Raster Data Processing Techniques

Giuseppe explained various aspects of working with raster data and image processing tools. He discussed file handling in GdalCalc, including handling multiband files and specifying band numbers, as well as the importance of checking data values and compression methods. Giuseppe also covered resampling techniques, explaining when to use different methods like nearest neighbor, bilinear, and cubic interpolation for categorical and continuous data. He emphasized the importance of compressing files to manage file sizes and explained how to use Digital Warp for projection changes and cropping based on coordinates or row/column numbers.

Data Processing and File Formats

The meeting focused on data file formats and processing tools. Giuseppe explained how to use the JDL command for data processing and emphasized the importance of specifying the correct data

type to avoid truncation or loss of data. He also discussed the GeoPackage format as an alternative to Shapefiles for storing large vector files, which can be read by OGR and used with GDAL. Giuseppe advised against building custom mosaics of satellite data, suggesting instead to use existing published datasets. The group also touched on the possibility of using Google Earth Engine with Python, though Giuseppe noted that he doesn't use it due to lack of transparency in the processing.

AI can make mistakes. Review for accuracy.

Please rate the accuracy of this summary.   👍 👎

Edit summary   Share

Thank you,

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

Zoom.com

55 Almaden Blvd
San Jose, CA 95113

+1.888.799.9666

© 2025 Zoom Communications, Inc.