Notes 2025-02-24

Carrie Hashimoto

Table of contents

Goals this week - SMAP-HB

- Fix spatial ref issue - solved
- Split up static and temporal data into zarr tiles - split up into nc tiles instead; done
- Train UNet with static data - not complete

Fix spatial ref issue

- Solved by letting reproject_match set the ref without using .write_crs
- Still create empty spatial_ref var when using rio.write_crs - have these leftover in tiles

Split up static and temporal data into zarr files

- Saved all temporal data in dynamic_predictors.zarr and target.zarr; originally tried saving 10 km and 30 m together, but kept getting dimensions from one dataset bleeding into the others
- When trying to convert static data to zarr, get warning that objects don't fit into zarr hierarchy - couldn't solve and didn't find much help online, so saved as nc instead

Split up temporal data

- Worked fine with just SMAP 30 m and 10 km
- Attempted solution: reproject_match IMERG to SMAP
- Realized IMERG has nv band and is missing spatial ref, so recropped and dropped time_bnds and added spatial refs
- Issue: having trouble avoiding getting extra 10 x 10 dimensions for 30 m resolution data

Software setup

- Have NVIDIA GeForce RTX 4060, which should be CUDA-compatible
- After first installation attempt, couldn't open NVIDIA app - need to investigate
- Try to install CUDA 12.8.0
- Pretty involved - recent issue with Visual Studio version not being compatible with CUDA; followed this tutorial to get Pytorch and Tensorflow
- Need visual studio: current version causes issue with CUDA installation, so custom download older version
- Tutorial pytorch installation recommedation didn't work bc only have versions 2+ and he used 1, so used from Pytorch website, below
- pip install torch==1.13.1+cu116 torchvision==0.14.1+cu116 torchaudio==0.13.1 –extra-index-url https://download.pytorch.org/whl/c
- pip install tensorflow==2.10.1
- Had to downgrade numpy

UNet prep

- Started data loader script

Goals this week - WRF-Hydro

- Finish training video - done
- Finish Jupyter Notebooks tutorial - need to fix ogr issue
- Set up ArcMap with WRF-Hydro GIS toolkit - done (using Pro)

Set up ArcMap with WRF-Hydro GIS toolkit

- issue: Rice only has Pro, not ArcMap
- Try with Pro
- They also have an open-source GIS tools beta version since ArcGIS is proprietary
- GEOGRID checking tool has issues but some other components work

Video content

- Create boundary shapefile
- Get variable from geogrid (export grid from geogrid file)
  - Using topography as an example
- Create routing stack: most basic input needed for gridded routing

Process geogrid - basic version

- skip forecast points;
- Input elevation raster;
- regridding – need to nest hydro (NWM - 250 m) into LSM (1000 m): factor of 4
- Number of routing grid cells to define stream: control density of stream network, use something close to NHD+ density (?) - 2 km2 -> 32 grid cells

– 2 sq. km per pixel * (250 m)^2 = 32 grid cells

- Naming: NED_30m_rf4_th32.zip for output (routing factor, threshold)
- This is supposed to create the basic routing stack: Fulldom_hires, GEOGRID_LDASOUT_Spatial_Metadata, GW-BASINS.nc, GWBUCKPARM.nc

More details for routing stack

- Process geogrid, adding .csv file with forecast points
- Option to create lake parameter file
- Then need to add reservoir shapefile/feature class - need lake shapefile
- Output will have same as before, but also LAKEPARM.nc, with record for each lake feature
- Failed, but log doesn't show any errors – haven't resolved this yet

Possible sources of error with lakes/forecast points processing

- Log file stops earlier than messages in ArcGIS, which say that the process completed successfully
- Maybe a saving issue, then

Examine outputs

- Examine outputs script makes a bunch of rasters
- Check flow accumulation - it worked! Even though the geogrid variable check script doesn't
- Check channel grid - output is different because I'm using the basic outputs and the tutorial uses the one including lakes… still looks alright though
- Check forecast points - same here; not included in basic output

- Stream order also worked
- Lake grid also not relevant with basic outputs
- Topography - also looks good :) based on input elevation raster, but converted to floating points and resampled to routing resolution

Alternative: WRF-Hydro Pytho GIS tools

- Try using the open-source GIS Python tools instead of ArcGIS toolbox? In beta, so might also fail

python Build_Routing_Stack.py -i geo_em.d01.nc -d NED_30m_Croton.tif -R 4 -t 20 -o croton_test.zip

- Create new conda env based on their instructions
- Ran without problems, but need to check documentation on how to include lakes/forecast points

To do this week

- Put combined tiles in RAPID
- Work on UNet development - meet with Adnan?
- Solve ogr thing and complete Docker tutorial
- Try example case pre-processing with current software?

Some small data things:

- Fix DEM naming - currently “band_data”
- Drop spatial_ref since it’s empty

Notes

- when you save the zarr, put the spatial ref into attributes, as argument in .to_zarr
- Run WRF-Hydro with inputs generated from GIS pre-processing