Notes 2025-03-03

Carrie Hashimoto

Table of contents

Goals this week - SMAP-HB

- Move data to RAPID

- Fix data to be saved in 360 x 360 chunks with zarr format

- Fix data to be saved with spatial ref as an attribute

- Train a basic unet

Goals this week - WRF-Hydro

- Try using pre-processing outputs to run WRF-Hydro example case in Docker

- Fix ogr issue in Docker

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" - did this and dropped "band" dimension
- Drop spatial_ref since it's empty: nope, can't do this or the rio.crs will disappear

Pytorch Tensorflow environment

- Need xarray to load data, so installing that too

For reference, originally installed the ML packages with pip, following this tutorial

- Created a conda environment using conda-forge with Python 3.10 in it
- 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

So additionally, pip install xarray==0.21.1 (older version of xarray to match numpy <2)

- Then also pip install netcdf4

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

Zarr software temporary fix for SMAP-HB

- mapping yaml: using the original version I created that specifies all versions works fine
- Recall from before, last time I tried something or other, it nevr solved, so sticking with the custom environment for now

Adnan meeting

- Adnan would like to just resize by chopping the dynamic data up into 360x360 chunks
- So the way I stored the data might not make sense for his purposes; rather, it would be easier to just reproject the coarse data beforehand and save them in one nc

WRF-Hydro

- To run WRF-Hydro with the outputs from the ArcGIS tutorial, pull a new version of the WRF-Hydro training 5.1.1 and bind mount the folder to it
- The example that I pulled before is 5.2.0, but since the WPS version is 5.1.1, see if it's possible to get 5.1.1 training to be consistent

docker pull wrfhydro/training:v5.2.0-rc1

docker run –name wrf-hydro-training -p 8888:8888 -v C:/Users/carri/Desktop/wrfhydro/docker_mount_example:/home/docker/mount -it wrfhydro/training:v5.2.0-rc1

WRF-Hydro

- In order to have the data to run it:
- WPS to create geogrid
- ArcGIS toolbox to create basic routing stack from geogrid and topography: Fulldom_hires, GEOGRID_LDASOUT_Spatial_Metadata, GWBASINS.nc, GWBUCKPARM.nc

This week, WRF-Hydro

- First run WPS tutorials again to verify geogrid I'm using is the one we DIYed
- WPS tutorials also create soil_properties.nc – here is where we need the updated R to proceed

To fix the stringr and ogr issues, start with default WPS docker

- First pull the docker training and bind the folder to it: docker run –name wrf-hydro-training -p 8888:8888 -v C:/Users/carri/Desktop/wrfhydro/docker_mount_example:/home/docker/mount -it wrfhydro/training:v5.2.0-rc1
- Then save the result as an image: docker commit wps-training-example-bindmount wps-training-example-bindmount:latest
- Then use dockerfile to make a new container with updated R: specifically need stringr, optparse, ncdf4, and plyr
- docker build –no-cache -t wrf_hydro_stringr .
- Tried adding libgeos-dev and libproj-dev

WPS issues

- Need to install stringr again: R scripts needed for preprocessing are create_wrfinput.R and create_soilproperties.R
- Create_Domain_Boundary_Shapefile.py requires ogr, which was the other stumbling block with getting the WPS training to work
- Got a Dockerfile that at least created an image without errors (wrf_hydro_stringr) – try running it

docker run –name wps-diy2 -p 8888:8888 -v C:/Users/carri/Desktop/wrfhydro/docker_mount_example:/home/docker/mount -it wrf_hydro_stringr:latest

WPS issues

- Stringr was successfully installed
- However, get this error when I try to run create_soilproperties.R:

Error in scan(file = file, what = what, sep = sep, quote = quote, dec = dec, : line 95 did not have 28 elements Calls: read.table -> scan Execution halted

- This happens when trying to read MPTABLE.TBL, which contains the parameters for Noah_MP
- Realized that I used the wrf-hydro image as the base, and it doesn't have geog_conus, the starting dataset
- So tried using the WPS image as a base instead, but need to modify the Dockerfile, maybe bc of differences in OS between the two: WPS image is v16, and WRF-Hydro is 20
- I think the only difference needed for the Dockerfile is libunistring0 vs 2

Yueyang's advice

- Install libiconv in Docker container
- Use pip to install gdal, and also install sf and rgdal

create_wrfinput.R vs real.exe

- real.exe takes inputs from realtime or reanalysis data, whereas create_wrfinput.R creates spatially uniform initial conditions

Meeting notes

- Check if you can include forecast points and not lake and make geogrid
- Try the one source tools instead
- Choosing version - need to use 5.3.0 because of impervious runoff scheme
- Goal: run example case with 5.3
- Or could artificially change soil moisture or Ksat to prevent infiltration
- Need to read about impervious runoff scheme
- For now just try using older versions with our data
  - Extra step before running model is to intersect NLCD impervious cover layer with soil layer and create a modified KSat layer based on this so that you have KSat = 0 there
  - How Noemi did it before: make porosity very small and then use transfer functions to recompute KSat based on that

Model

- COuld keep 360x360, a bit bigger than the paper
- multiple channels x 10 - getting heavy
- COuld do PCA analysis on static data to reduce dimensionality from 10 to 5-6
- Mask for missing data adds another layer
- FOr predictors, do nearest neighbor interpolation
- Report on how well the model is doing: RMSE, R2, etc.