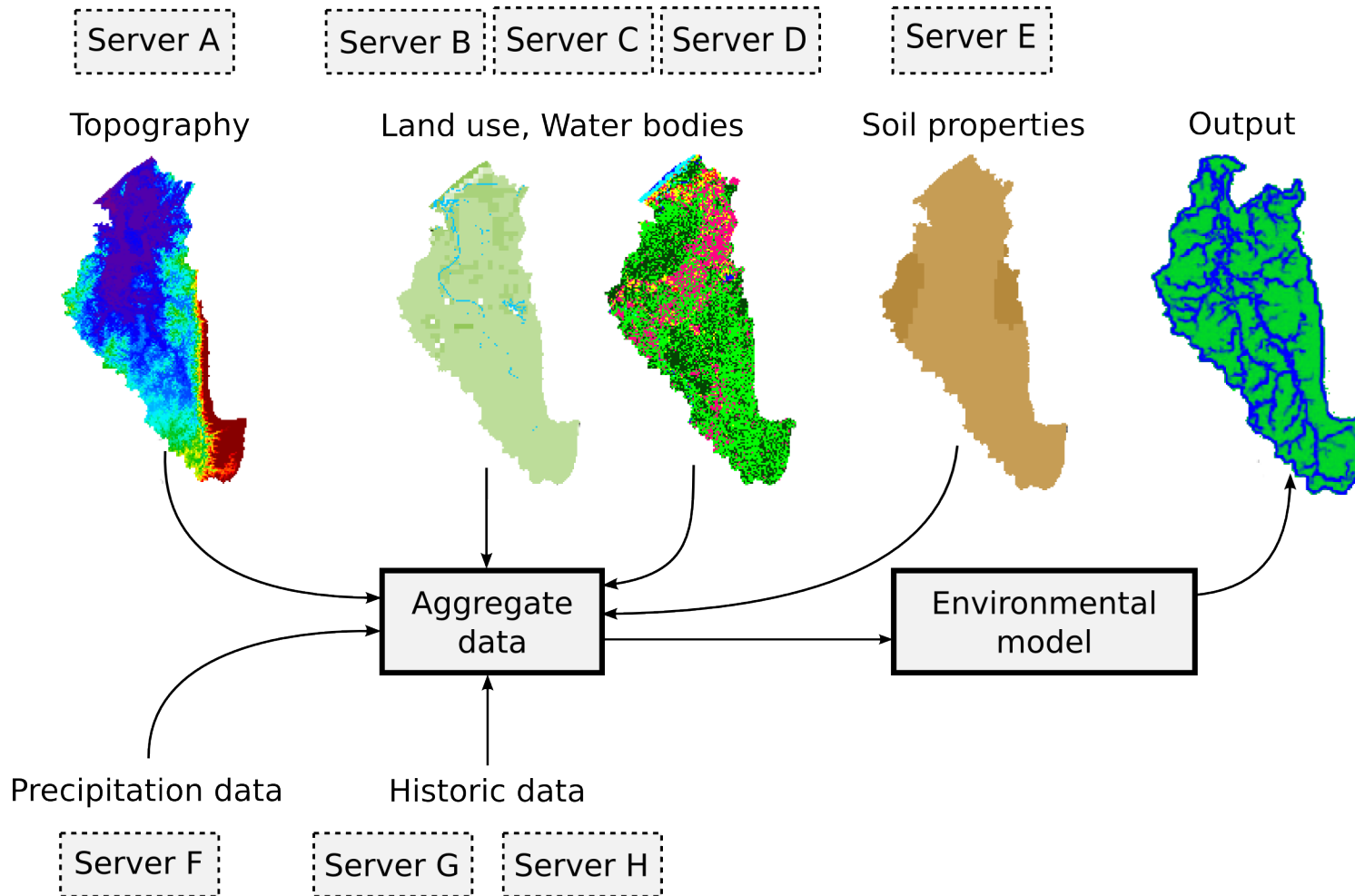


A Python-Based Automatic Data Aggregation Framework for Hydrology Models

Lucas C. Villa Real, Frank Liu and Thomas Osiecki
IBM Research

The problem



DomainBuilder

Automates acquisition of **public data** given a lat/lon bounding box

- Various communications protocols

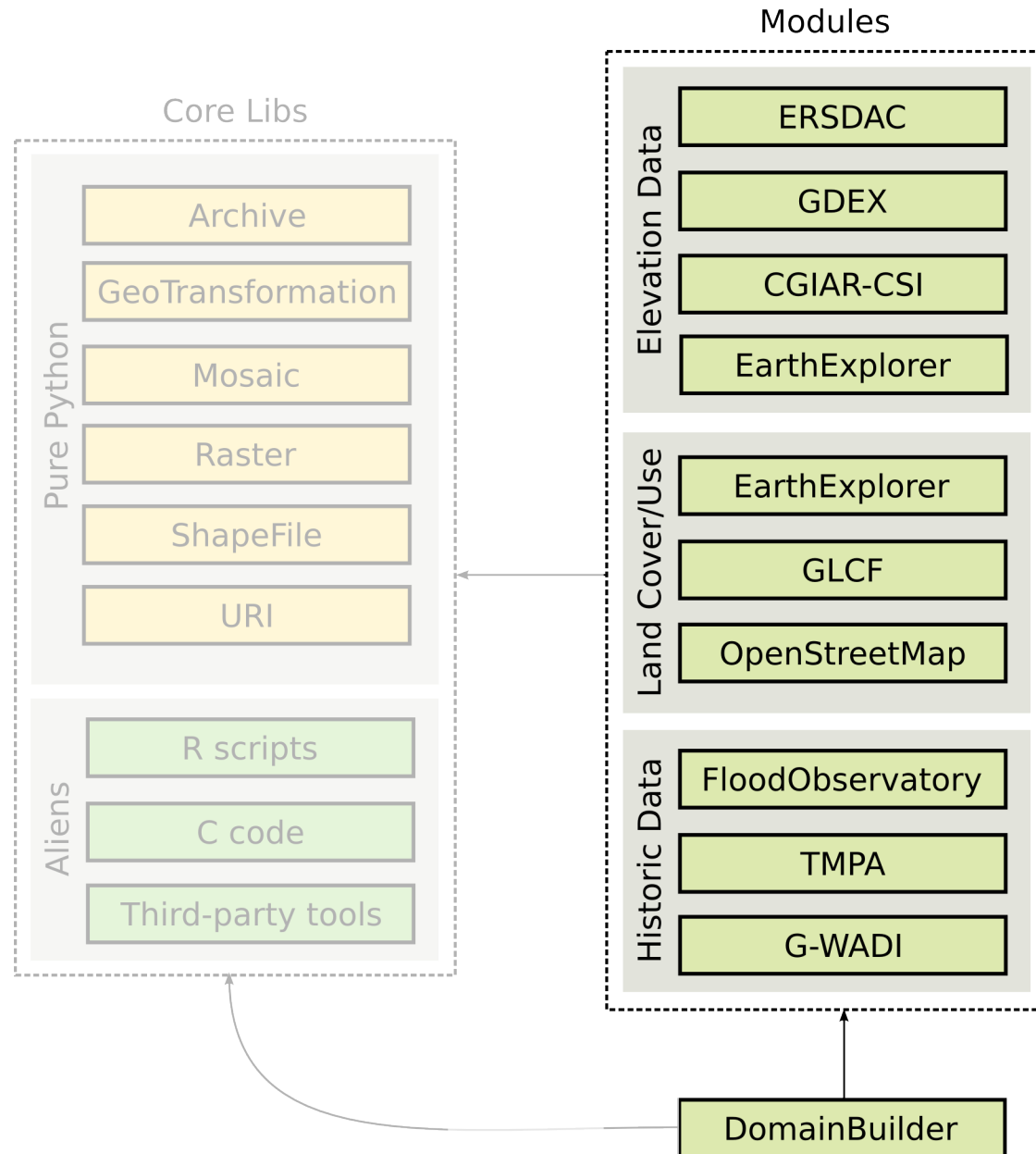
Performs **data transformation** on the fly

- Vector and raster formats

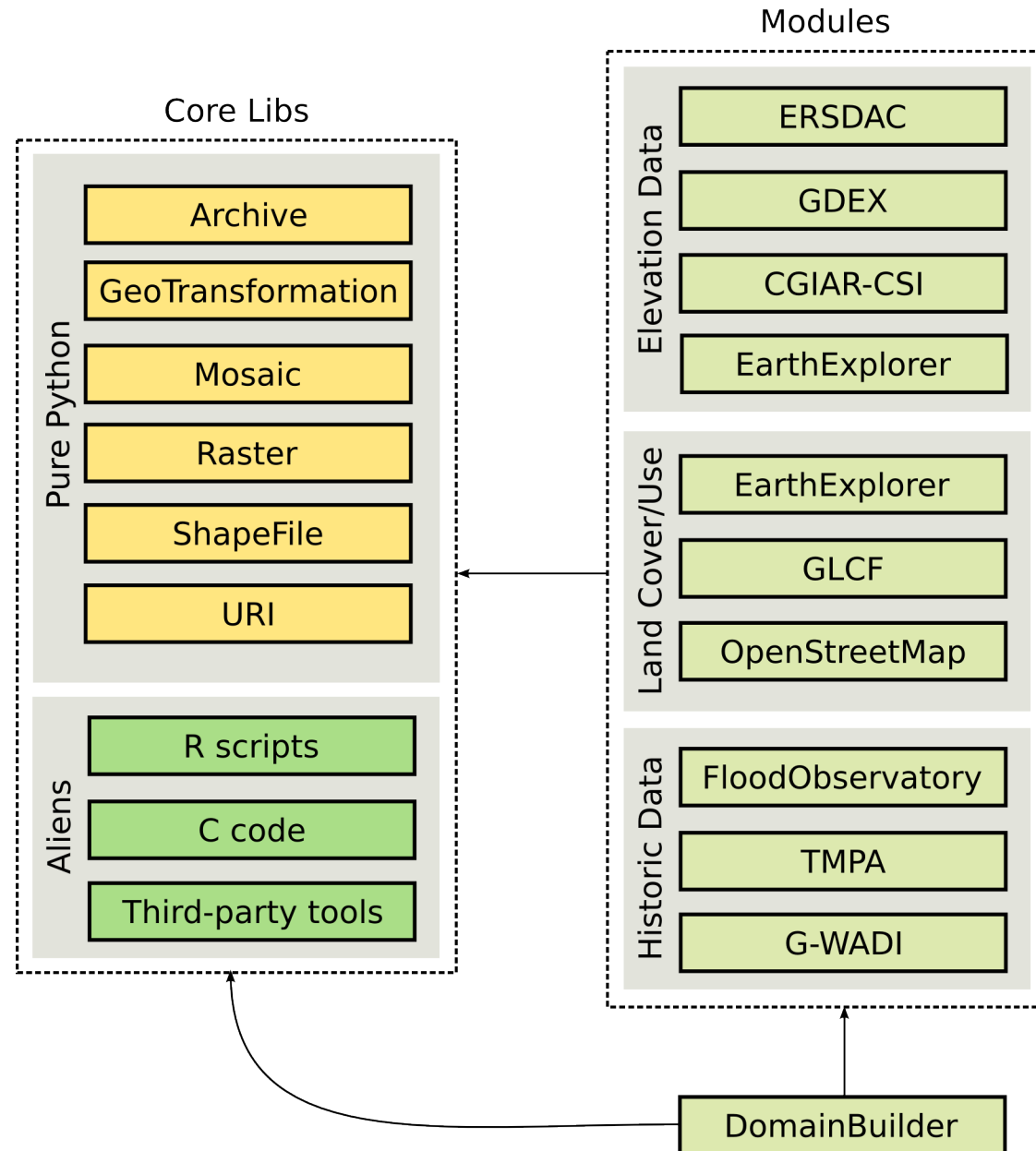
Outputs cropped / scaled / rotated files

- Same number of pixels and spatially aligned

DomainBuilder



DomainBuilder



DomainBuilder

```
EarthExplorer_Server = "https://earthexplorer.usgs.gov"
EarthExplorer_Settings = { "login": "username" "password": "secret" }
EarthExplorer_Datasets = [
    "EMODIS NDVI",
    "L8 OLI/TIRS"
]

SRTM_Servers = (
    "http://droppr.org/srtm/v4.1/6_6x5_TIFs/",
    "ftp://srtm.csi.cgiar.org/SRTMv41/SRTM_Data_GeoTiff/"
)

...
```

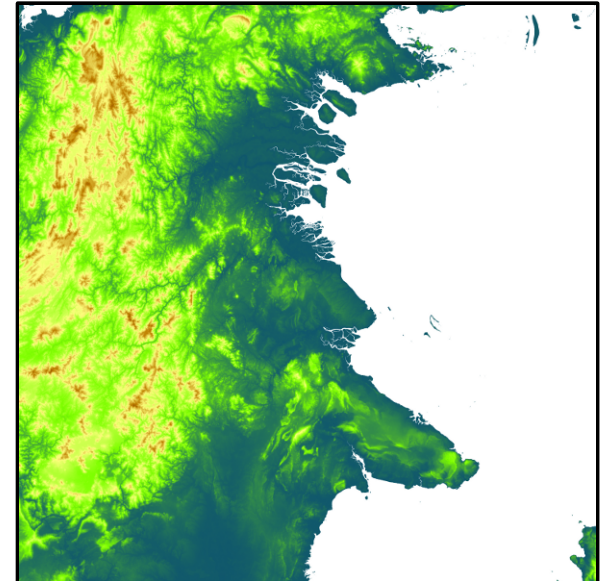
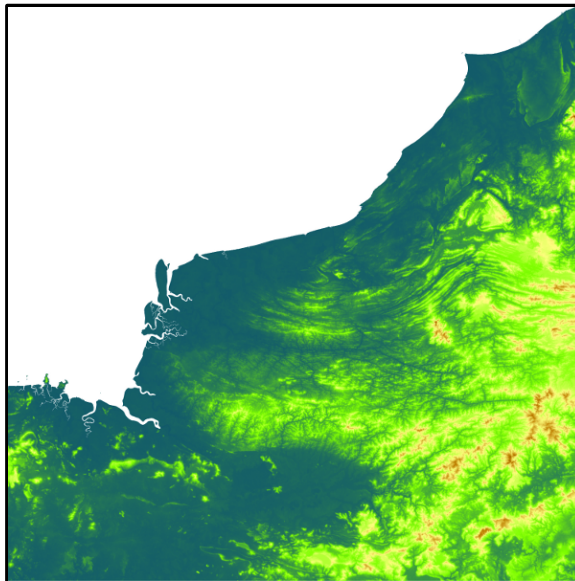
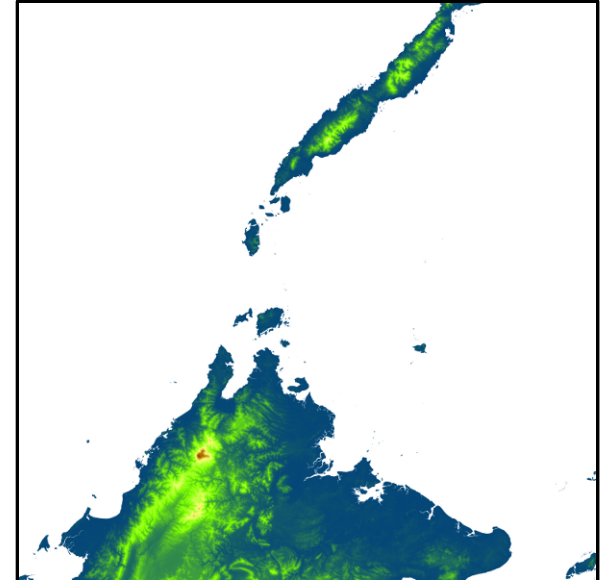
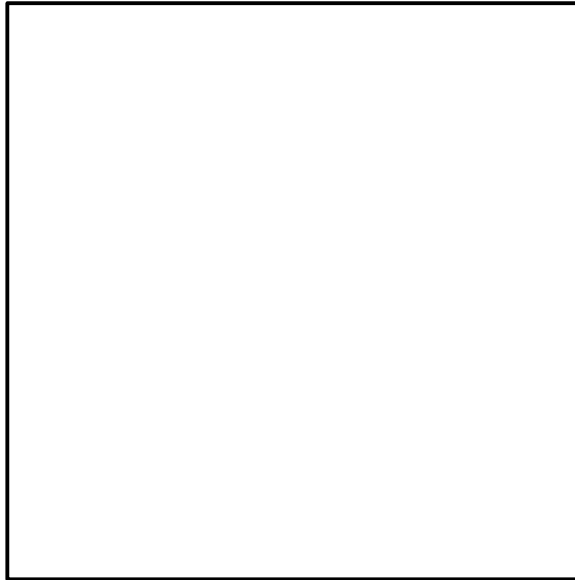
Elevation data

Tile processing

- GDAL

Watershed delineation

- SAGA GIS
- NumPy mmap



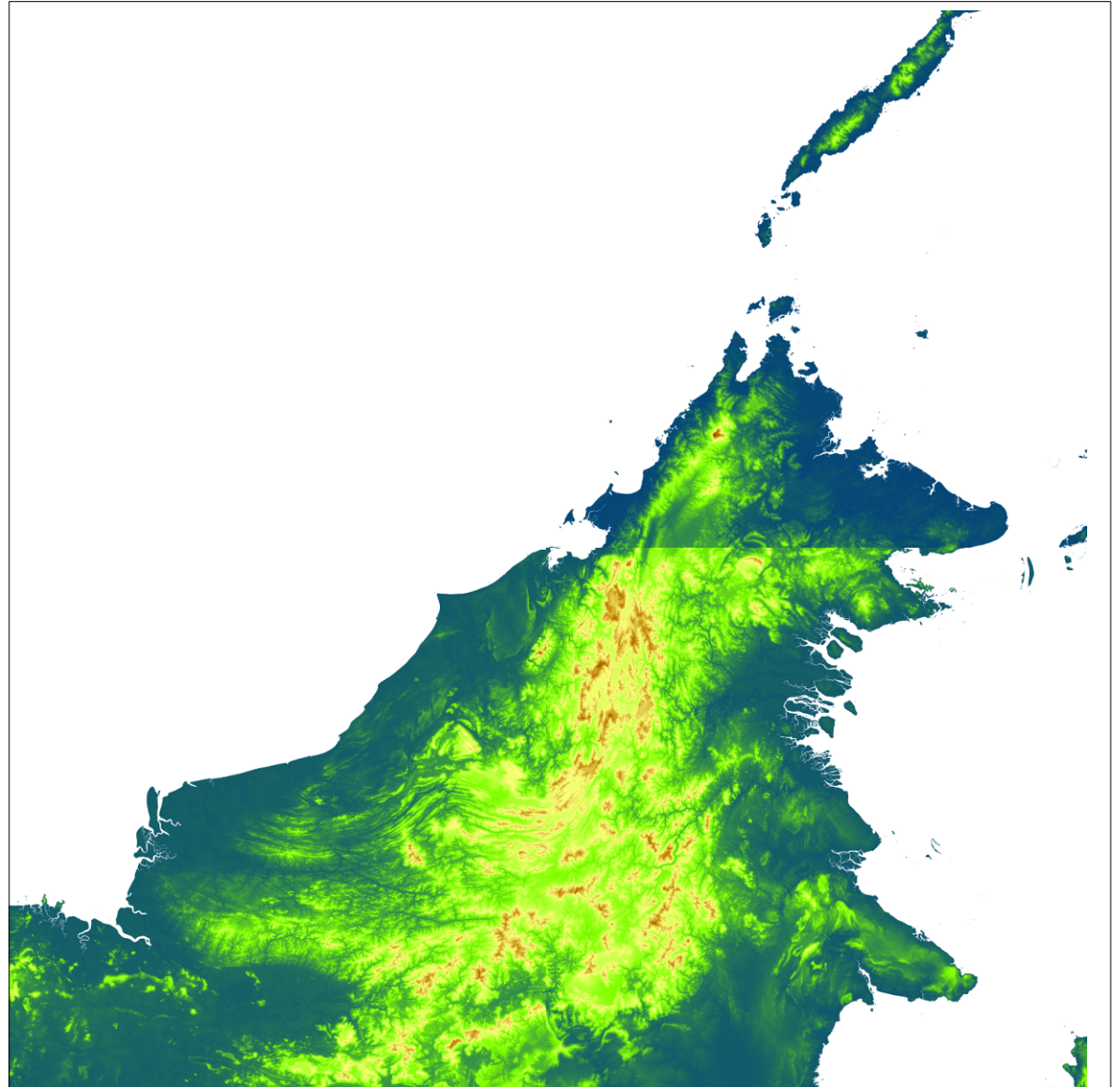
Elevation data

Tile processing

- GDAL

Watershed delineation

- SAGA GIS
- NumPy mmap



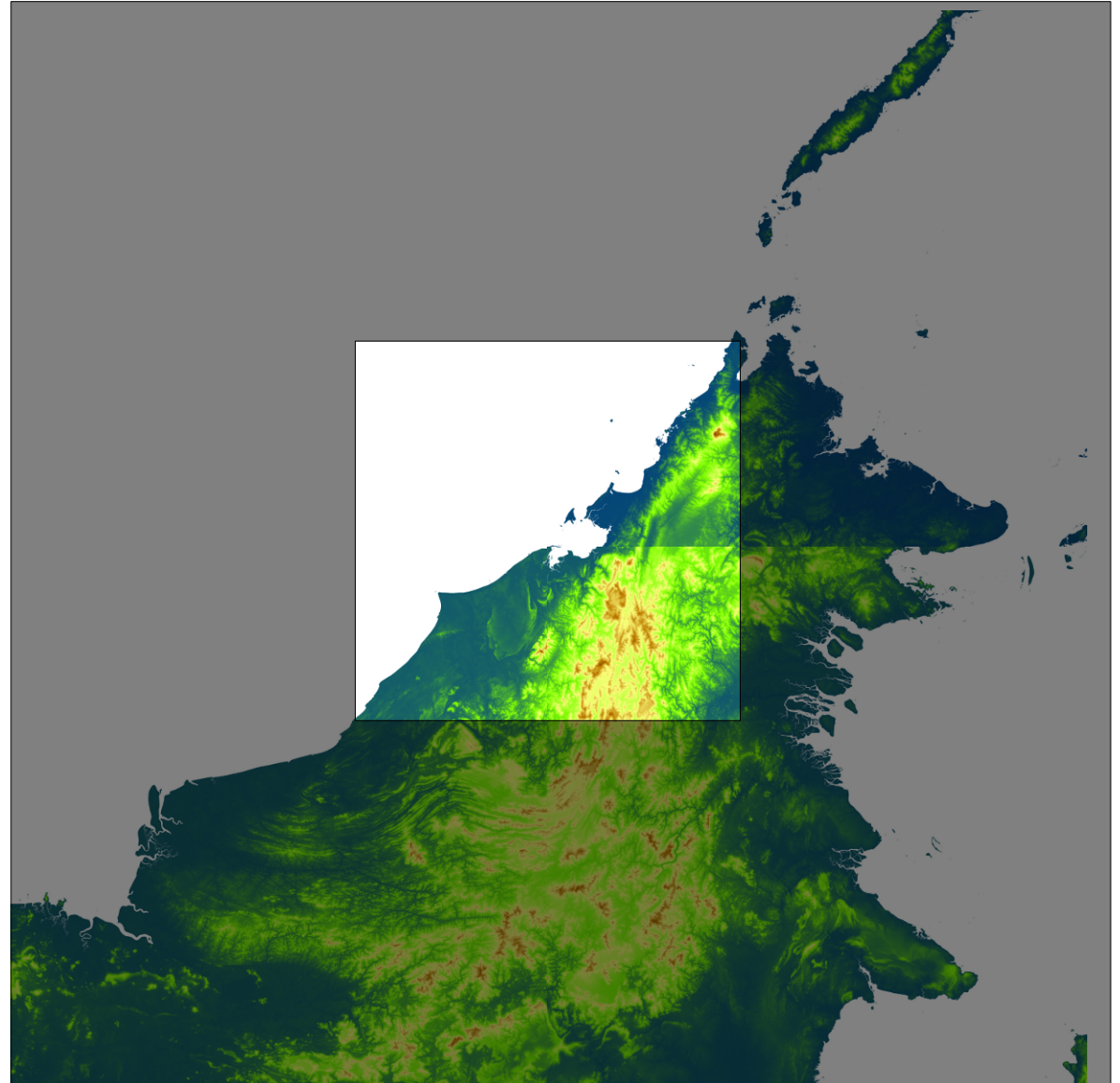
Elevation data

Tile processing

- GDAL

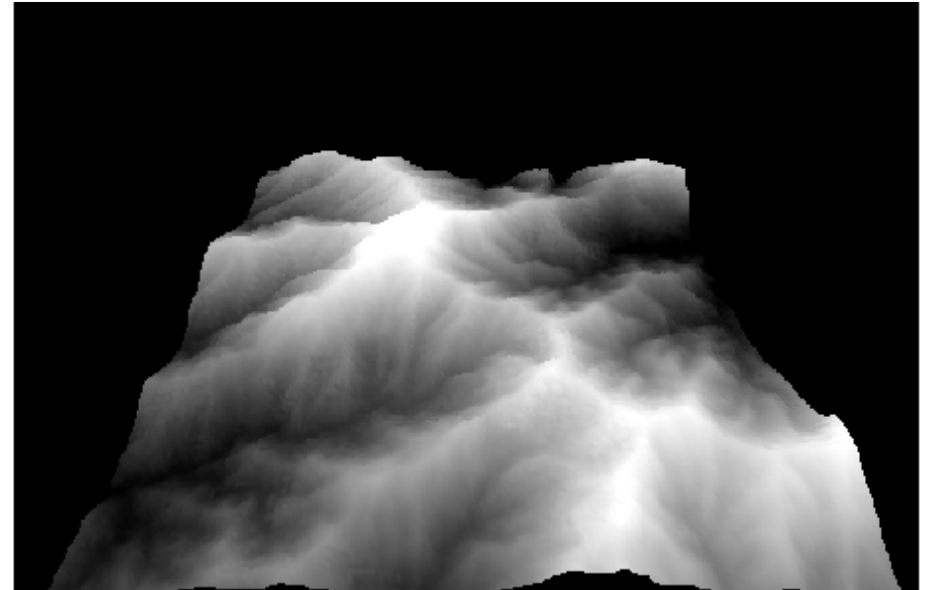
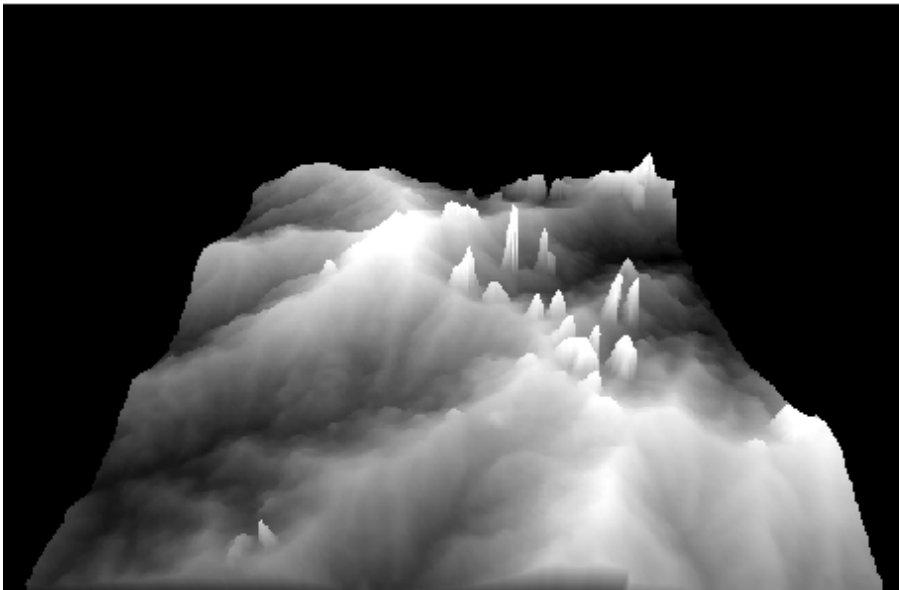
Watershed delineation

- SAGA GIS
- NumPy mmap



ASTER GDEM post-processing

Vertical artifacts removal using SRTM as reference



Open Street Map

Water bodies, land use

OSM data:

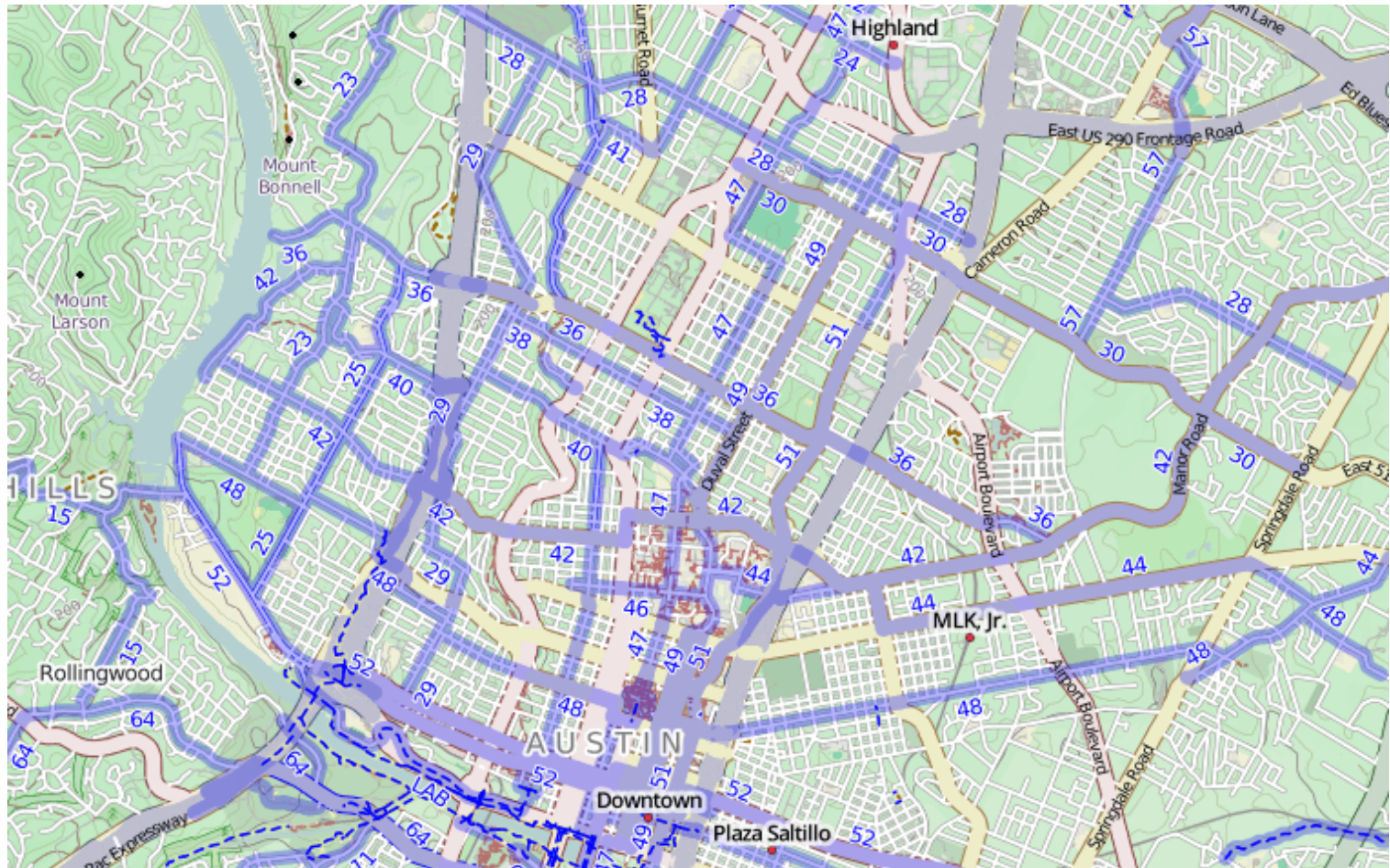
- Overpass API

OSM to SVG

- MapWeaver

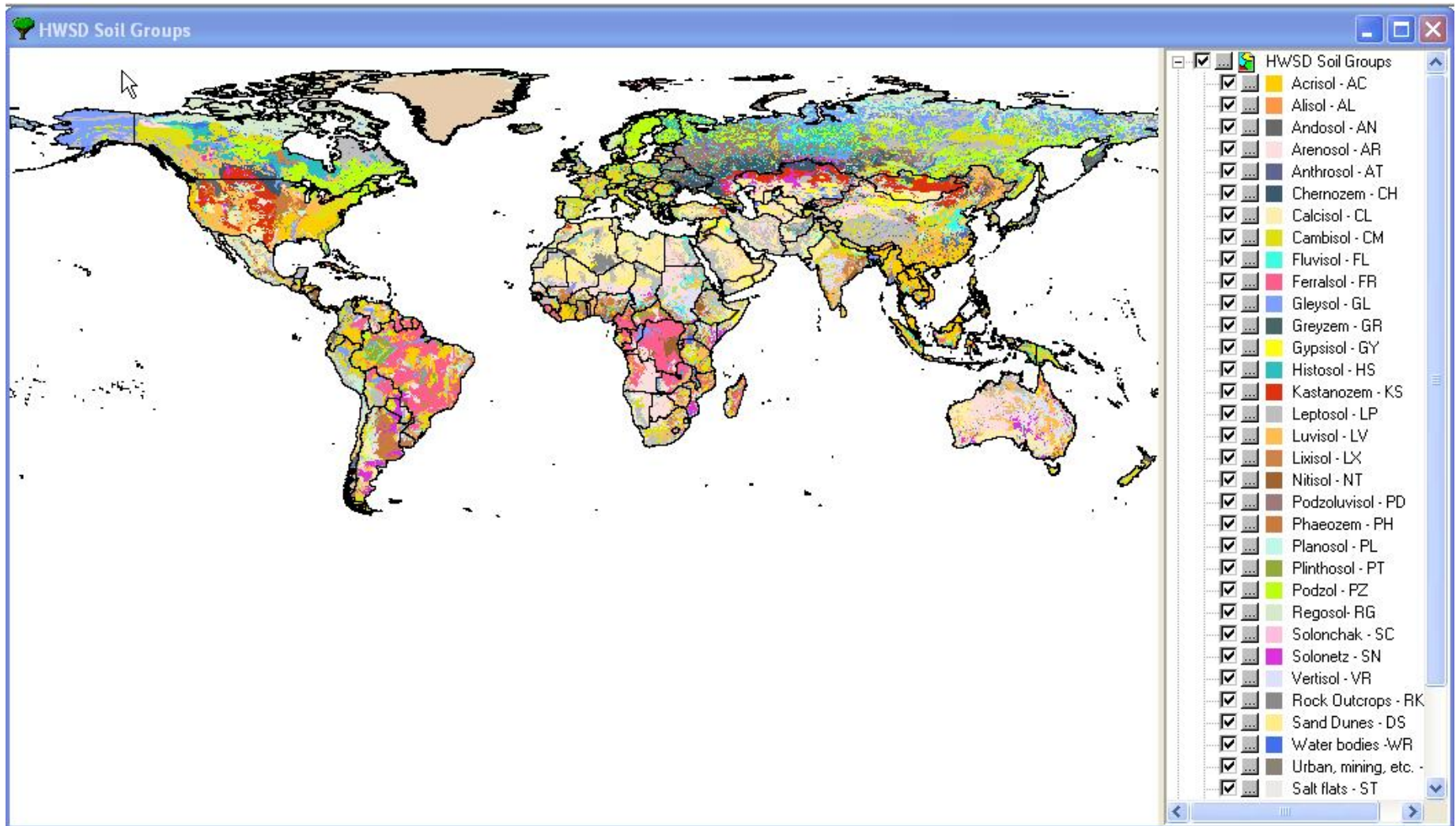
SVG to Raster

- ImageMagick



Soil Data

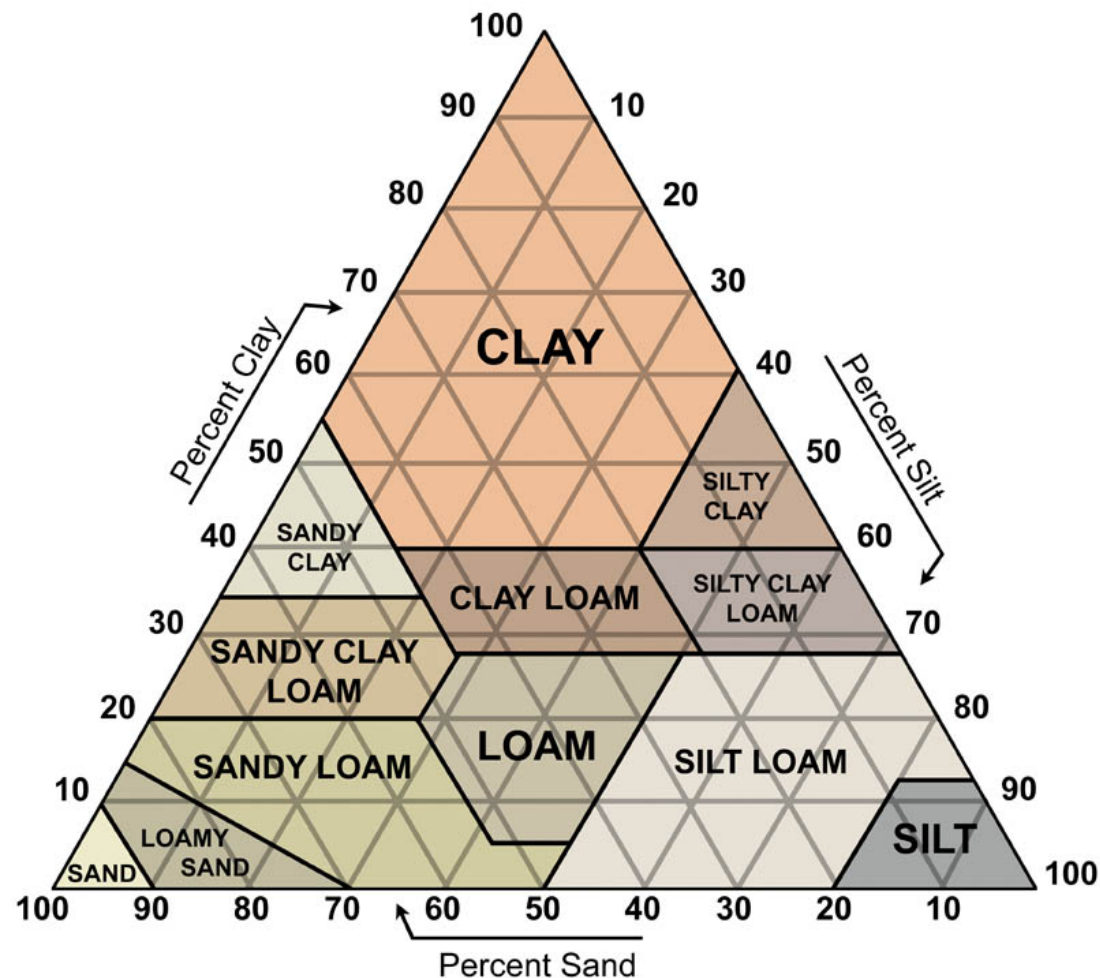
Uses MDB-Tools to extract soil properties from Harmonized World Soil Database



Soil Data

Soil properties are used to classify the soil

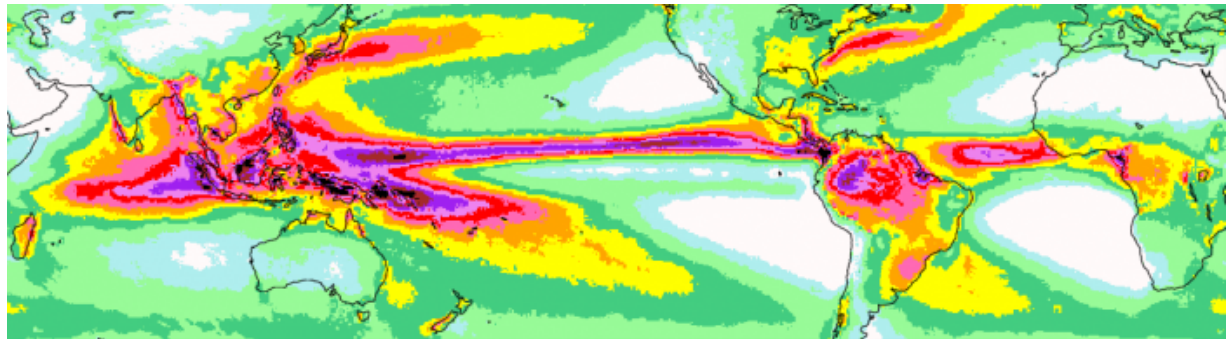
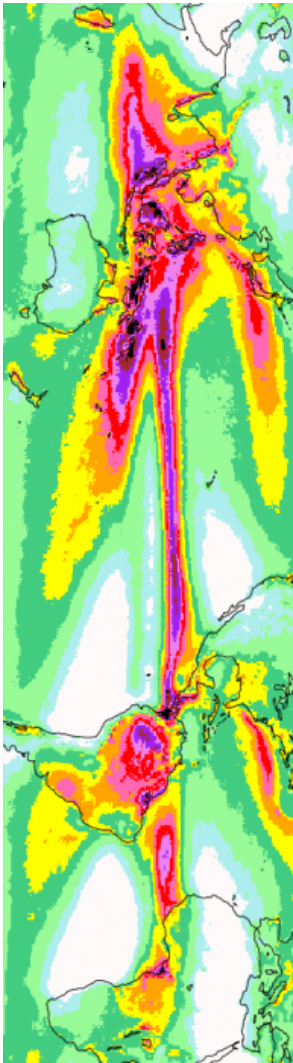
Outputs a matrix with the same resolution as the DEM



Historic precipitation data

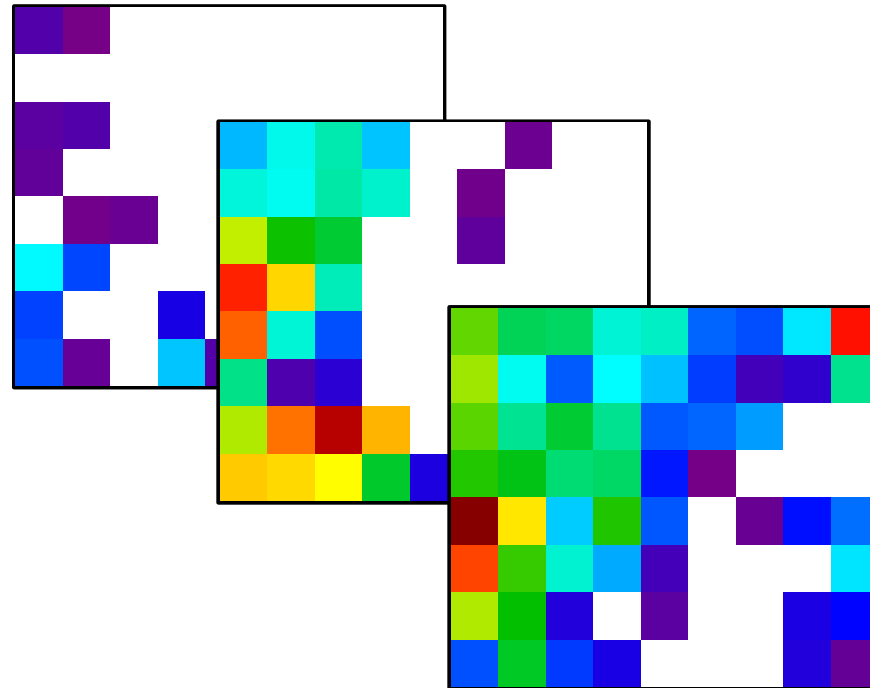
TMPA:

- Takes starting and ending dates as input
- Rotates and crops as needed
- Python multiprocessing library



Extreme Precipitation

Takes historic TMPA data as input, outputs series of NetCDF matrices



Conclusions

- From prototype to stable code very fast
- Not only an integration language
- Very pleasant to use

A Python-Based Automatic Data Aggregation Framework for Hydrology Models

Lucas C. Villa Real, Frank Liu and Thomas Osiecki
IBM Research