Introduction

Large terrain datasets provide the foundation for the work carried out by the British Geological Survey (BGS) in developing various environmental data products.

It is vital that:
1. datasets are kept up-to-date
2. any uncertainty associated with the dataset is effectively accounted for and communicated

Hurdles hindering these considerations relate to:
- available memory
- processing time

Presented is our workflow for creating derivative datasets from a digital terrain model (DTM) considering uncertainty.

Methodology

- Fully automated Python workflow working on Geotiffs.
- Essentially convolves a 8x8 window across the dataset to derive slope and aspect values along with associated uncertainty (cf. Heuvelink et al., 1989).
- Data are tiled into 10 km areas according to the British National Grid.
- Tile indexing enables fast searching and partitioning of neighbouring GeoTiff tiles to deal with calculations at grid corners.
- Calculations at each tile are fully vectorised (numpy).
- Uncertainty simulations require the breaking down of each tile into manageable blocks to meet memory requirements.
- Outputs aspect/slope grids with associated uncertainty.

Results and delivery

- Provision of derivative datasets and uncertainty surfaces to be integrated into the BGS product development workflow.
- Minimizes data that need to be held in memory: reducing memory requirements and processing time.
- Increases ability to re-deploy as required to incorporate data updates.

Conclusions and next steps

- Moving window operator can be adjusted as required (just pass the function e.g. for roughness etc.)
- Provides the BGS with a ready-to-go uncertainty simulator.
- Uncertainty products will be fully incorporated into all future products and associated updates.

Examples of existing approaches for storing and processing geospatial data

- Once you have lots of tiles and you want to store, query and process them more efficiently, it’s worth investing time in integrating them into some type of framework.
- Below are some examples of available software which can assist with various use cases, helping with databases, visualisation and processing:
  - Proprietary
    - GIS databases
    - Oracle Spatial
  - Open source
    - GDAL
    - PostGIS (no longer updating raster support)
    - Open Data Cube
    - Cloud Optimised Geotiffs (COG)
    - Prosto Raster

Data structures

- The speed-up achieved by working with your data once integrated into one of the above (other) frameworks is based on the underlying architecture.
- There are 2 core architectures for spatial data:
  - Raster: Quad-tree
  - Vector: R-tree
- New architectures are being developed to further increase efficiency e.g. K²-tree (Brisaboa et al., 2017).

What’s a quadtree?

- Spatial area of interest
- Binary representation of extent
- Raster block break down
- Block quadtree

References