Bespoke 4th Year Dissertation Proposal

Fin Gray

Project Overview

Create meaningfully accurate bare-earth version of a specified region. Done using custom data sanitisation pipelines leveraging open source data to feed a machine learning pipeline. This ML pipeline will interpolate out man-made disruptions to the original environment. The output will be a version of the given area, free from human impact, in an easily visualisable format with commercial and academic merit.

Broad Pipeline

1. Using open data extensively, create pipelines to transform into a ubiquitous standard.
2. Using this standard, manually mask areas of extensive human interference.
3. Train a machine learning model to automatically mask additional areas creating a new dataset.
4. Train separate models to interpolate different aspects of terrain make-up and deformation. Finally using a GAN to combine, contextualised by expert opinion.

Broad Timeline

Term begins

literature review concluded.
Pipeline established for major data formats (likely elevation & national survey).

ML iteration

Iterate on prototypes, training models, resolving issues.
Output file type confirmed.
Dissertation started.

Submission & Release

Submit dissertation and project.
Package & publish project to GitHub & others.

ML model prototypes

Rough prototypes of each ML pipeline completed.
Compute resources secured.
Manual masking begun.

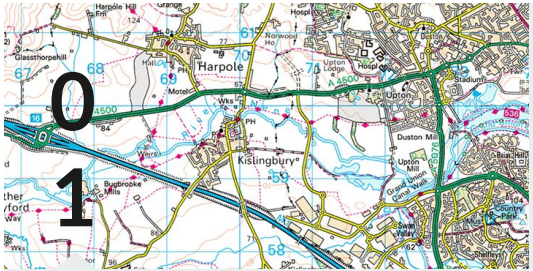
Wrap up

Finalise models.
Finalise pipeline.
Complete dissertation drafts 1&2.

Stretch Goals

- Global
- Searchable
- Generative AI outputs

Additional Information

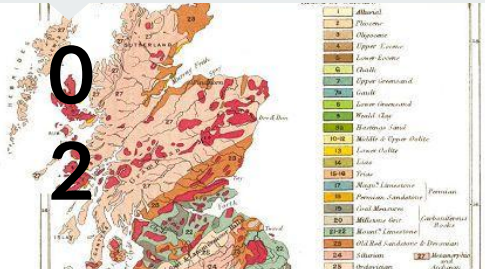


Reliance on Geoscience

With the aim to be as representative as possible opinions on meaningful inputs (geological data, etc.) and critiques on outputs would be invaluable

Initial data sources

- Free OS OpenData Maps
  - Natural Earth, OpenStreetMap, EarthExplorer, etc.
- Global DEM data sources
  - SRTM30 plus / SRTM15 plus, etc.

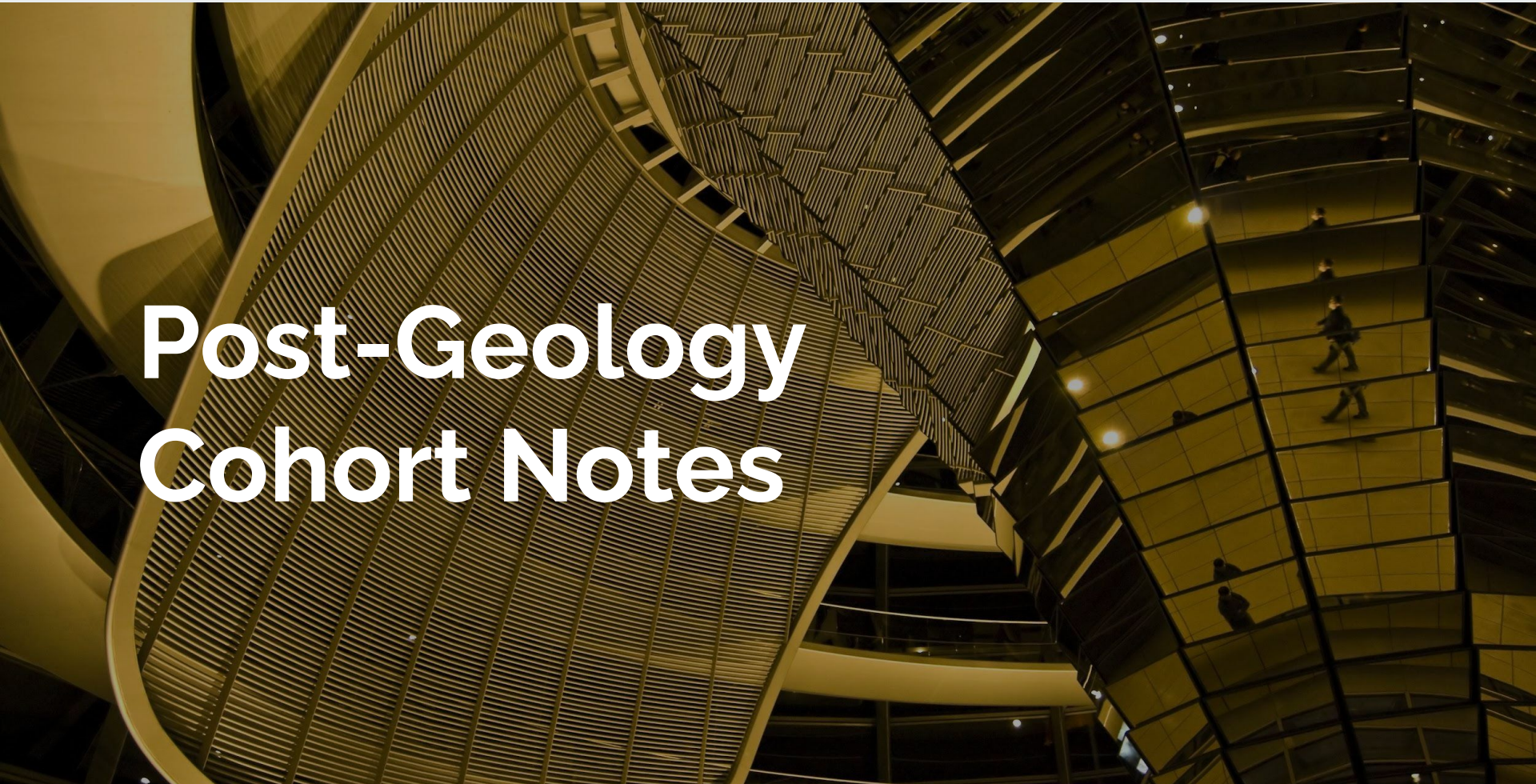


Support

Gerardo Camarasa - SoCS advisor *(tentative)*
Ed Parsons (Google Earth) - Informal advisor

Project objective

Using data science and machine learning, create a pipeline to generate a bare-earth digital twin using open and easily accessible data



Post-Geology Cohort Notes

Suggestion: City size over time

Using the development of cities since Ordnance Survey's beginnings (1745), we can get several examples of city sprawl over time to help train the model to help determine the pre-development characteristic.

Cumbernauld and other such new cities and towns will be valuable.

Suggestion: Coverage

Divide the specified region into a grid, selecting focus points in each to begin manual masking to get better generalisability. Gradually increasing as the model improves.

This will also eliminate potentially difficult areas for the masking model to evaluate.

Suggestion: Human Impact Score

The likelihood of areas needing to change can be tied to the proximity to inhabited areas. Using heatmaps for things like population density etc. we can form a proxy for the human impact score to enhance the model.

Why not planetary?

This question was raised during the meeting, the answer was it has not been considered.

Despite this major components of the pipelines created could be reused for this purpose.

Why what is the focus?

Creating meaningfully accurate 3D environments

Reasoning: To provide crucial assets for PhD and further simulation work.

What resolution?

As much as the procured compute allows.

Initially starting with 30m elevation data.

What about dams and rivers?

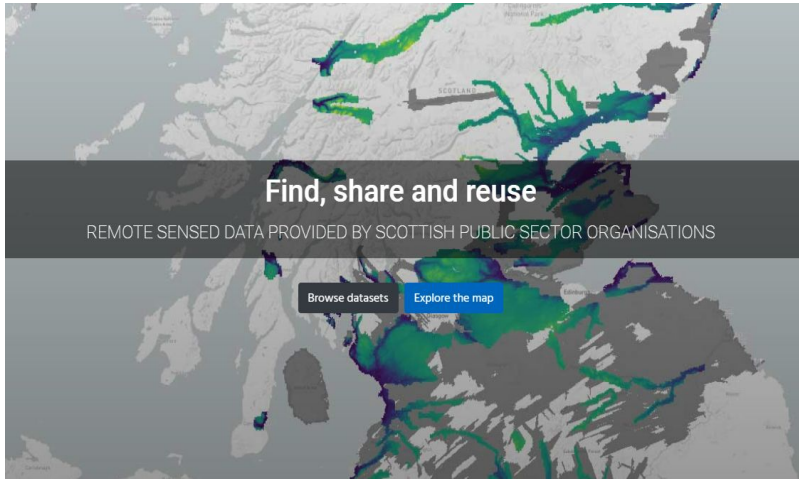
This was a level of complexity that was not considered. An may dramatically change results.

Potential solutions providing that it is not captured by the model: Have a bespoke model explicitly for this. Train it on rivers around that word that have data before and after a dam or significant human or animal interaction.

Additional Resources #1

Scottish Remote Sensing Portal

One of the geologists mentioned this resource. Lidar information from the major valleys in Scotland. The majority of the information being around inhabited areas.



https://remotesensingdata.gov.scot/

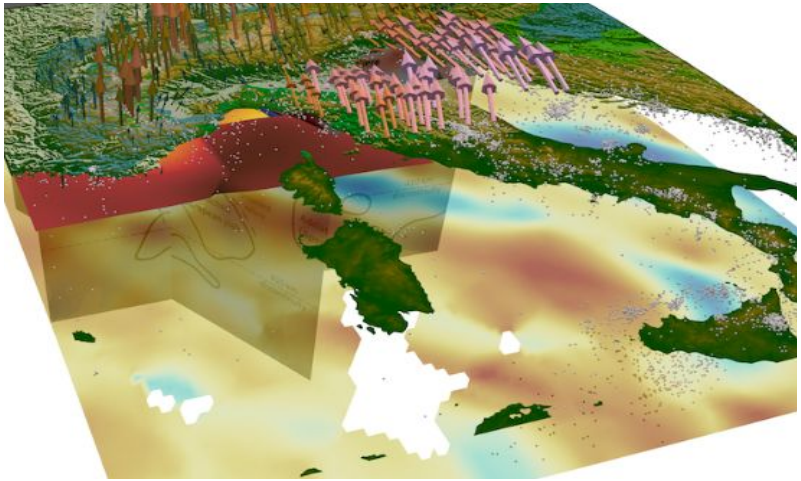
Additional Resources #2

GeophysicalModelGenerator.jl

Another geologist suggested this resource which may speed up the initial stages of the pipeline.

"The aim of this package is to help with this, by providing a number of routines to easily import data and create a consistent 3D visualisation from it in the VTK-toolkit format"

https://github.com/JuliaGeodynamics/GeophysicalModelGenerator.jl



SoCS Supervisor

Dr. Gerardo Aragon Camarasa
Senior Lecturer (Computing Science)

Gerardo.AragonCamarasa@glasgow.ac.uk

Gerardo's Notes

Ensure open source

Email Gerardo when call for custom projects announced.

Mentioned Gaussian Splats for the ubiquitous standard

Mentioned started with a lower resolution - 100m/1km

Mentioned rushing a full MVP to know which parts of the pipeline

Compute gets busy around February, rush prototype so not in a deadlock

Suggestion: Glasgow flyover

Purely a thought on a final deliverable.

A clean fly over of Glasgow rendered from the 3D asset that changes from video, to open street view, and has the man made elements regress through time -if an option- or fade into uninhabited version.