

Group Project : Earth and Environmental Science

Data Integration and Visualization for the Low Burn Hall Estate - Datacube

Huixiaqing Liu - fhnp96

1 Introduction

Effective management of natural resources requires a comprehensive understanding of the target environment. This often needs the integration of data from multiple sources to characterize the site of interest.

This project aimed to develop a datacube incorporating field-collected resistivity and ground-penetrating radar (GPR) data, satellite imagery and MODIS-derived products for the Low Burn Hall Estate. The original objective was to enable interactive 3D visualization of the datacube. However, due to computational limitations, a 2D composite image has been created as a substitution.

This report details the data collection, processing and integration methodologies, presents the results, discusses the limitations and proposes future improvements.

2 Data Collection

2.1 Fieldwork

Resistivity data was collected at two locations (P1 and P2) within the estate using a resistivity meter. Measurements were taken at increasing electrode spacings to sample increasing depths.

Ground-penetrating radar (GPR) surveys were conducted using a Pulse EKKO system at positions P1 and P2. An additional GPR survey was carried out at position P1 using the wide-angle reflection and refraction (WARR) method. GPR two-way travel times were converted to depth using estimated velocities.

Potential 1/2 Spacing	Current 1/2 Spacing	Measure 1	Measure 2	Measure 3	Measure 4
5	15	78.1	78.1	78.2	78.9
5	20	42	41.9	41.8	41.7

Table 1: Example resistivity input data

Tool	Interpretation	GPR Line	Position (m)	Depth (m)	Velocity
Point	Pink	Lineset/line1	0.43	0.67	Moist Soil (0.1 m/ns)
Point	Pink	Lineset/line1	4.03	3.96	Moist Soil (0.1 m/ns)

Table 2: Example GPR interpretation input data

2.2 Provided Data

Satellite imagery of the estate was provided as a georeferenced TIFF file.

MODIS land surface temperature (LST) and vegetation index products were provided as CSV files at a 1 km spatial resolution for 2001 and 2019.

Year	Day	Easting	Northing	LST_day	LST_night	NDVI	EVI	Easting_km	Northing_km
2001	0	426625	538600	0	287.52	0.7755	0.5446	426	538
2001	1	426625	538600	300.78	285.94	0.7755	0.5446	426	538

Table 3: Example MODIS input data

3 Data Processing and Integration

3.1 Resistivity Data

The resistivity data was cleaned by removing negative and outlier measurements. The cleaned data was then inverted using a 6-layer model in pygimli to estimate the true resistivity structure at each location. Figure 1 shows the raw and inverted resistivity data for Position 1, while Figure 2 shows the same for Position 2.

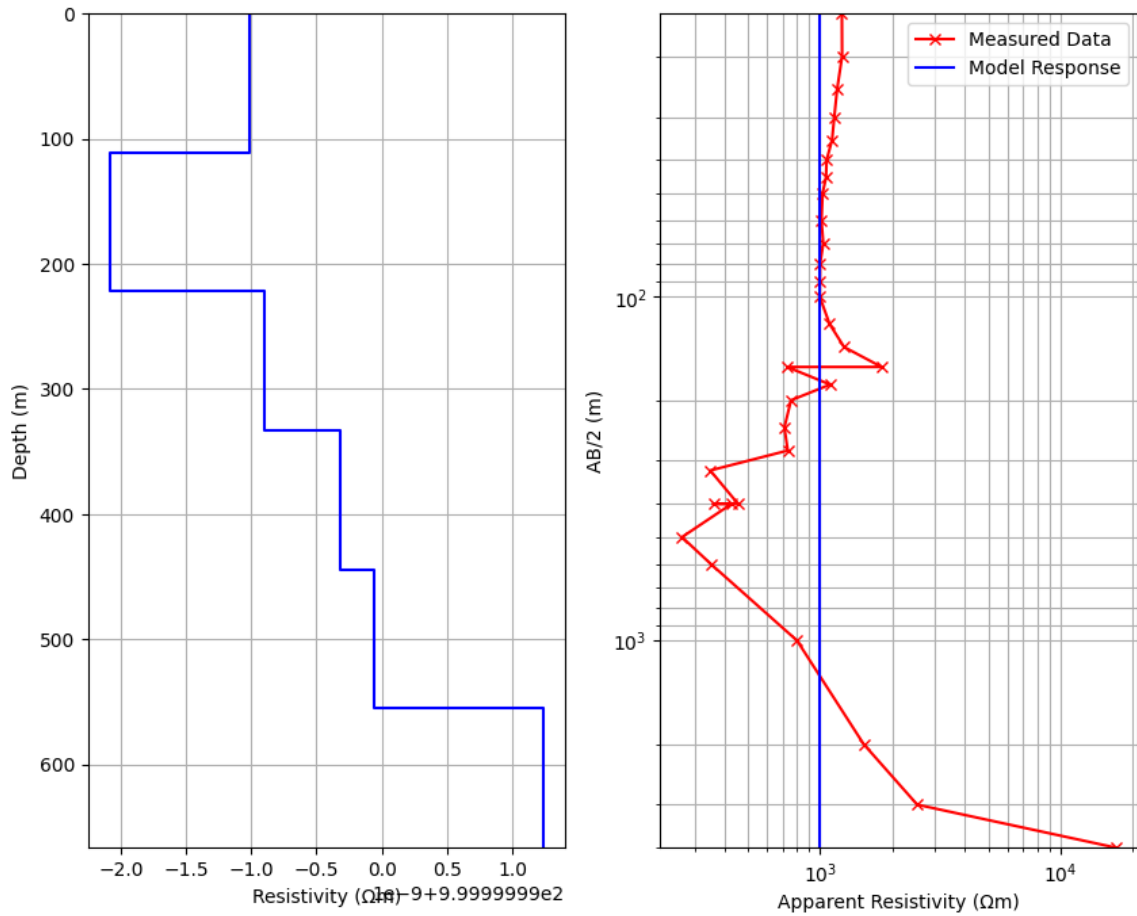


Figure 1: Inverted and raw resistivity measurements at P1

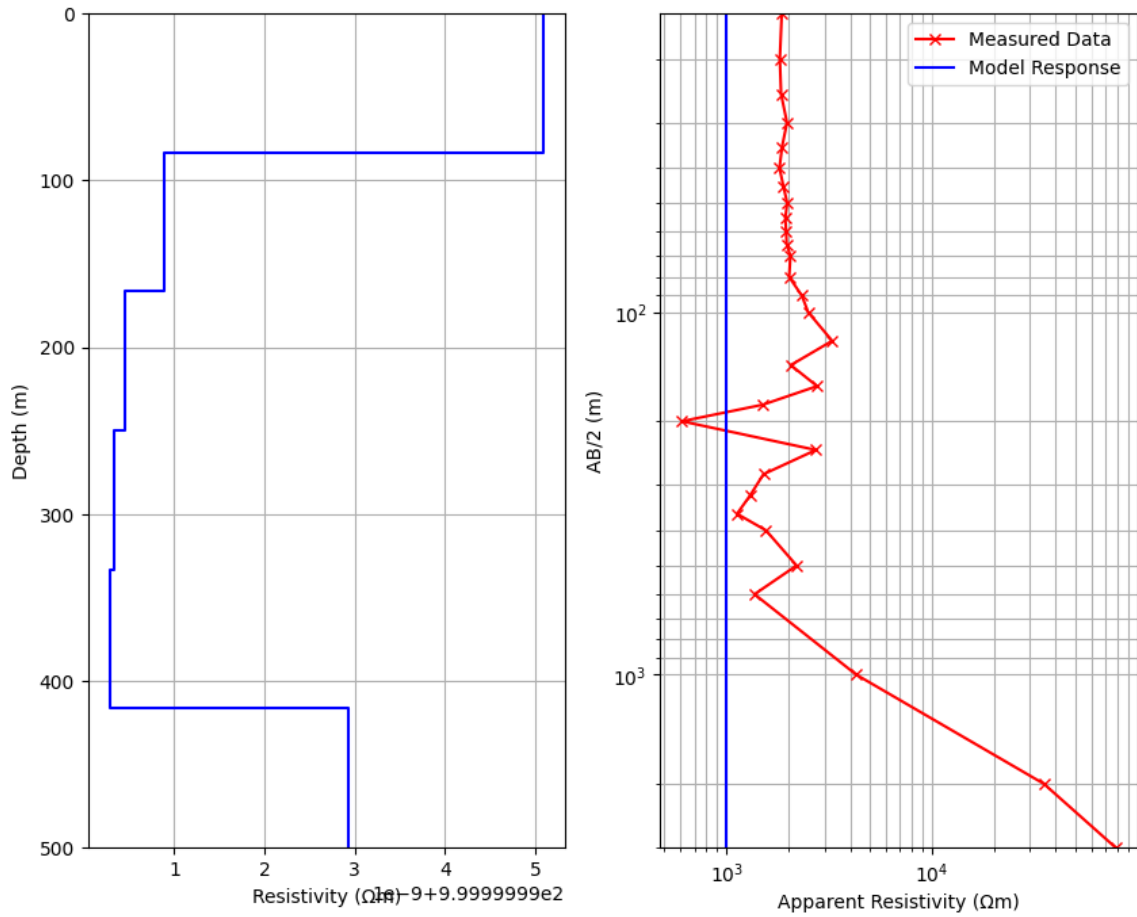


Figure 2: Inverted and raw resistivity measurements at P1

3.2 GPR Data

The raw data was stored in .gpz files, which is a proprietary format used by the EKKO Project software. Due to licensing restrictions, we did not have access to the EKKO Project software, which limited our ability to export and analyse the GPR data directly.

As a result, we relied on the provided GPR interpretation points for our analysis. These interpretation points lacked spatial reference information, so we manually assigned the locations to the resistivity positions P1 and P2 based on the survey notes. The GPR reflector depths were used directly without further processing.

3.3 Satellite and MODIS Data

The satellite imagery was resampled to a common grid using bilinear interpolation.

The MODIS data was cleaned to remove duplicates and gap-filled to a regular space-time grid using xarray. The LST day, LST night, NDVI, and EVI layers were extracted and resampled to match the satellite image grid.

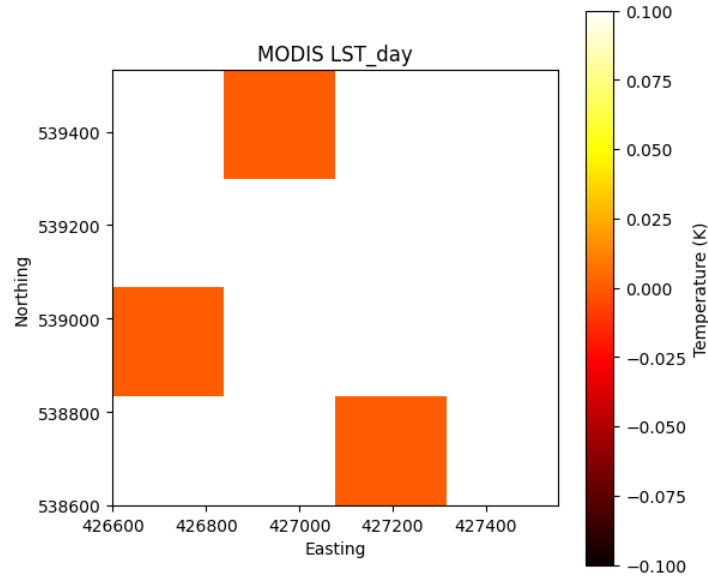


Figure 3: MODIS LST Day

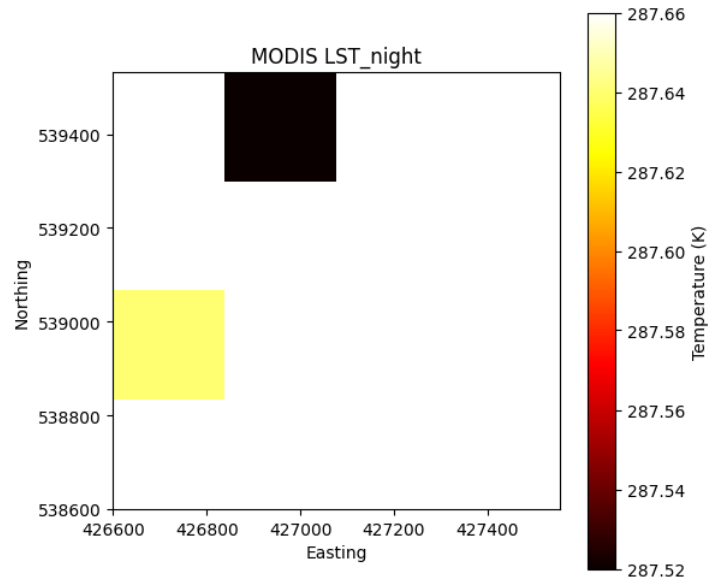


Figure 4: MODIS LST Night

3.4 Data Integration

The processed resistivity, GPR, satellite, and MODIS layers were combined into a single xarray Dataset object. This represents the datacube, with dimensions of easting, northing, and depth/layer as appropriate for each data type.

4 Visualization

The datacube layers were visualized as a 2D composite image using matplotlib (see Figure 5). The satellite imagery forms the base layer, with MODIS LST day values overlaid using a colormap and transparency. The resistivity models and GPR interpretation points are plotted at their respective locations.

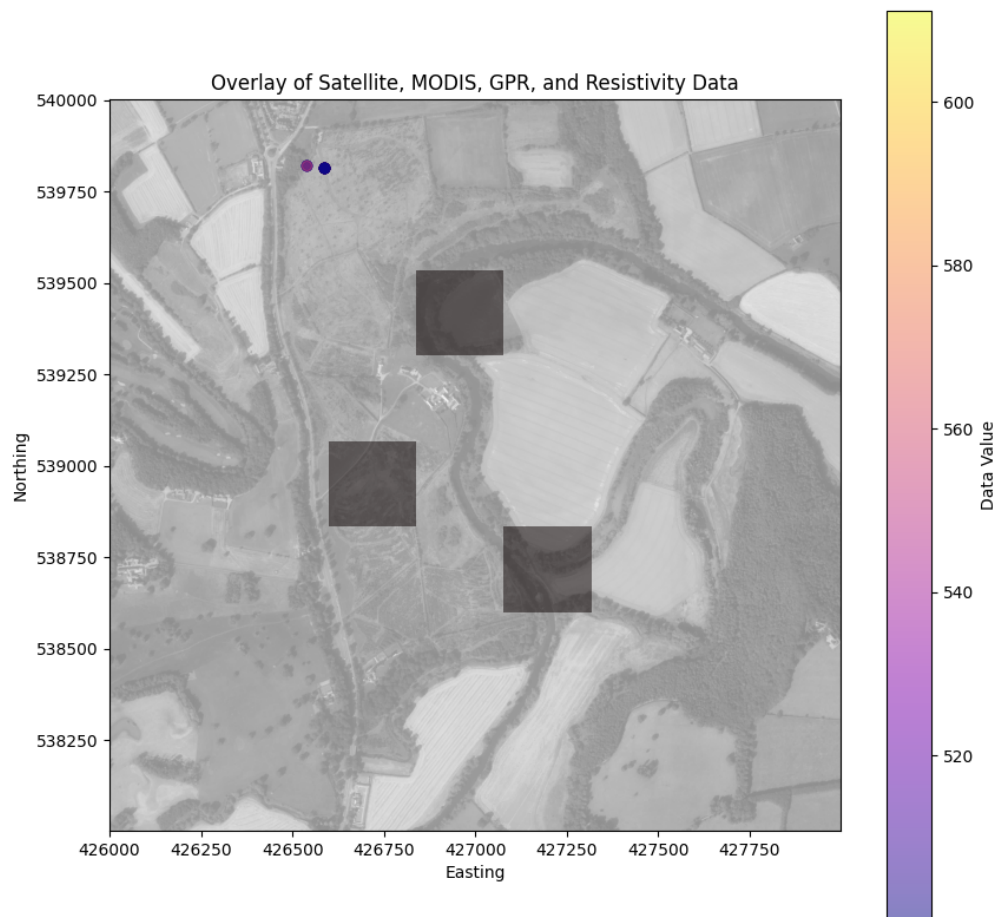


Figure 5: Composite 2D image of the data layers.

A 3D interactive plot was attempted using plotly, but the large size of the dataset caused memory issues. (see Figure 6)

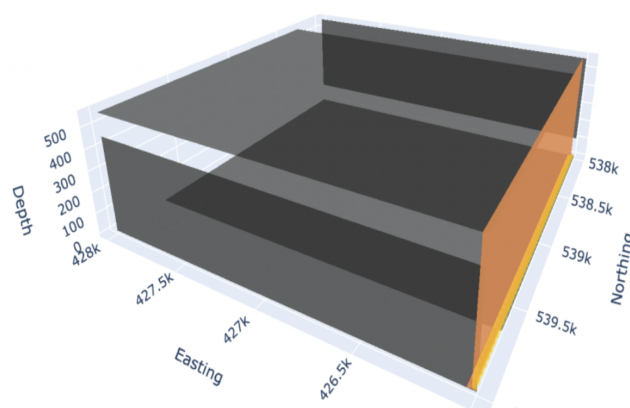


Figure 6: Faulty 3D image

5 Discussion

5.1 Achievements

Despite the challenges encountered, the project achieved a successful integration of field measurements, satellite imagery and modelled data products into a 2D datacube structure. This required careful data cleaning, re-formatting and spatial alignment. The resulting 2D composite visualization provides an informative overview of the spatial patterns and interrelationships of surface and subsurface properties across the Low Burn Hall Estate.

The combination of inversions of field data (resistivity), direct measurements (GPR), remotely sensed imagery and modelled products (MODIS) leverages their strengths. The 2D image communicates a comprehensive characterization of the site that would be difficult to determine from any individual data source.

5.2 Challenges and Limitations

The lack of spatial reference information for the GPR and resistivity data was a significant challenge. The manual assignment of these data to approximate locations based on survey notes introduces potential inaccuracies in the integrated product. Improved field data collection practices that include recording GPS locations at each measurement site would improve data integration and enhance the reliability of the datacube.

The computational demands of processing and visualizing the large, multi-dimensional dataset posed difficulties. In particular, memory constraints prevented the realization of the planned interactive 3D datacube visualization. This limitation could potentially be overcome by employing cloud computing resources or high-performance computing infrastructure to handle the intensive data processing and rendering.

The MODIS data products, while useful for capturing broad landscape patterns, have a coarse spatial resolution of 1 km. This limits their utility for characterizing finer-scale spatial characterization within the estate. Incorporation of higher resolution remotely sensed data (e.g., Landsat, Sentinel) could provide more detailed insights.

5.3 Future Directions

Following methods could be taken to enhance the datacube and its derived products.

- Improved field data collection protocols that include GPS locations for all measurements
- Leveraging high-performance computing resources to enable 3D interactive datacube visualization
- Incorporation of higher resolution remotely sensed data products
- Application of advanced machine learning algorithms (e.g., Deep Learning) for data fusion and pattern extraction
- Development of a web-based interface for easy access to and interaction with the datacube
- Adoption of the Open Data Cube framework for enhanced data management and analysis

The Open Data Cube (ODC) framework is a powerful tool for managing and analysing large volumes of geospatial data. It provides a set of APIs and tools for indexing, querying and processing data stored in a datacube format. The ODC framework offers several advantages that could significantly enhance the Low Burn Hall Estate datacube.

- **Scalability:** The ODC framework is designed to handle PB-scale datasets, making it well-suited for managing the growing volume of data collected at the estate.

- **Interoperability:** The framework supports a wide range of data formats and sources, facilitating the integration of diverse datasets into a unified datacube.
- **Analysis-ready data:** ODC provides tools for preprocessing and standardizing data, ensuring that it is analysis-ready and consistent across time and space.
- **Efficient data access:** The indexing and querying capabilities enable rapid extraction of relevant subsets of data for analysis.
- **Collaborative development:** ODC is an open-source project with a growing community of developers and users, providing opportunities for collaboration and continuous improvement of the platform.

6 Conclusion

The integration of multi-source spatial data for the Low Burn Hall Estate shows the potential of datacubes for comprehensive environmental characterization. By combining field measurements, remotely sensed imagery and modelled data products, we can gain a general understanding of the interrelationships between surface and sub-surface properties.

The key challenges encountered in this project - particularly the lack of spatial reference information for field data and the computational demands of large, multi-dimensional datasets - highlight areas for methodological improvement. Future work should prioritize the collection of GPS locations for all field measurements and the utilization of high-performance computing resources for data processing and visualization.

7 GitHub Repository

The code and datasets used in this project are available in the following GitHub repository: [DU-EE-Term_2-Datacube](#).