Interdisciplinary Project in Data Science

Pfennigbauer Johannes, 11902046

March 26, 2025

Project Summary

Title: The Impact of Subsurface Scattering on Microwave-Derived Soil Moisture Retrievals

Supervisor: Dr Martin Schobben, Dept. of Geodesy and Geoinformation, TU Vienna

Co-Supervisor: Dr Laura Vana Gür, Dept. of Statistics and Mathematical Methods in Economics, TU Vienna

Start of Project: July 2024

Domain-Specific Lecture: 120.110 Data Retrieval in Earth Observation

Abstract

Microwave remote sensing is a widely used method for estimating soil moisture, but its accuracy in arid regions is reduced due to subsurface scattering, where microwaves penetrate dry soil and are reflected from underlying hard surfaces. This study investigates rock fragments and distinct horizons in soil as drivers of this anomaly using additional environmental data, including elevation, slope, land surface temperature and precipitation. Linear models identified interactions between coarse fragments and other environmental variables as key factors influencing subsurface scattering, aligning with existing research. However, these models struggled to capture the complexity of non-linear interactions and provided limited predictive accuracy. Non-linear models, particularly Random Forest, slightly improved prediction accuracy but were harder to interpret. Despite their differences, both modeling approaches highlighted silt content and land surface temperature as crucial interacting variables for coarse fragments. These results highlight the need for advanced explainability techniques, greater temporal granularity in the data and climate-zone specific fine-tuned models and to improve the accuracy of soil moisture prediction in challenging environments.

1 Introduction

Since 2010, soil moisture has been considered an essential climate variable in the Global Climate Observing System. It is known to be a reliable predictor of floods and droughts, an essential parameter for agricultural management, and a meaningful estimate of other climate variables. [1]
A common technique to measure soil moisture content on a large scale is microwave remote sensing, which uses microwave backscatter signals as an estimate of soil moisture. This technique is based on the premise that the backscatter signal increases with the soil moisture content due to the change of the dielectric contrast at the soil-air interface in wet soils. However, under certain circumstances such as frozen soils, this technique does not deliver reliable results. Similarly, an anomalous behaviour was found when microwaves are reflected from dry soil, especially in arid regions. This effect, called subsurface scattering, occurs when the microwaves pass through the first few centimetres of dry soil and get reflected by hard surfaces beneath, as illustrated in Figure 1. This phenomenon results in an inverse relationship between soil moisture and the backscatter signal compared to the typical assumption. [2]
As a consequence the data obtained using the traditional assumption would show high soil moisture content in these arid regions, where in situ measurements prove that it is actually very dry. Therefore, the remote sensing group of Vienna University of Technology has developed a method to detect such anomalies on radar images using a reference soil moisture dataset. The analysis shows that the soil type and the coarse fraction of the soil could be reliable variables for the detection of this type of anomaly. [3] [4]
This project aims to investigate the influence of rock fragments and distinct horizons on the described anomaly using auxiliary data such as terrain elevation, soil group and soil composition on the example of Australia. Therefore the research question is formulated as follows:

> Are rock fragments and distinct horizons in soil drivers of the subsurface backscattering anomaly in dry, arid regions?

2 Data

The occurrence of subsurface scattering is obtained by the correlation of Sentinel-1 backscatter timeseries and reference soil moisture data from ERA5Land provided by the European Center for Medium-Range Weather Forecasts (ECMWF) at 9 km resolution. The data is provided by the remote sensing group of the Vienna University of Technology and covers the whole of Australia, including various climatic zones from wet at the coast and dry in the interiors, as pictured in Figure 2.

The provided dataset contains relative soil moisture content (`sm`), as the median soil moisture content of ERA5Land over the same time range as used for calculating the subsurface backscattering signal (`rsub`), and the 5^{th} (`sm005`) and 95^{th} (`sm095`) percentile respec-

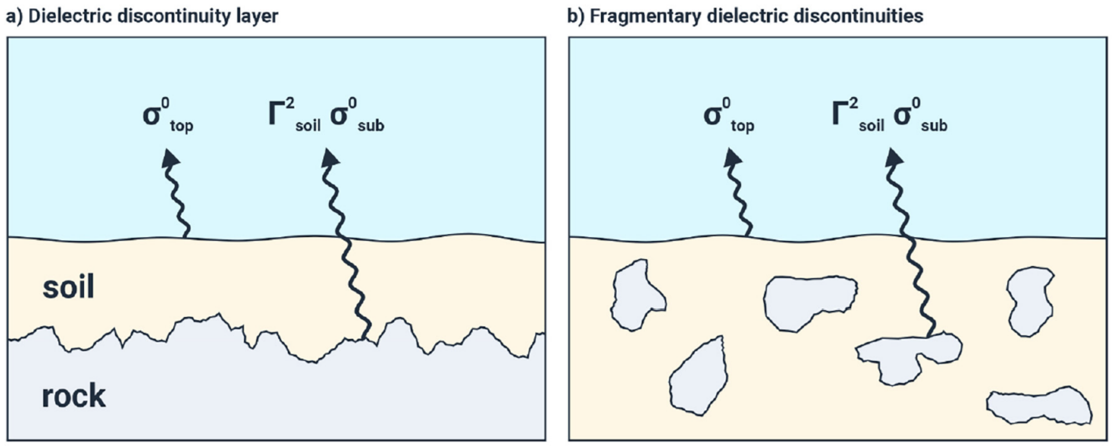

Figure 1: Illustration of subsurface scattering from [2]. Rock surfaces beneath the top soil layer (a) and dispersed stones (b) are the main causes of the phenomenon.

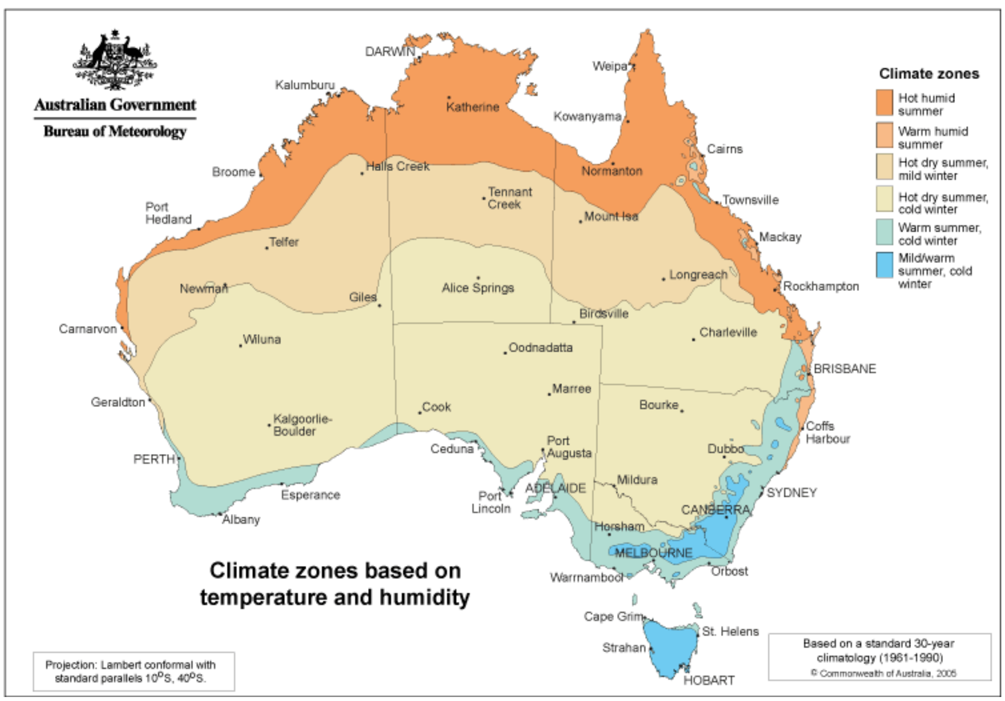

Figure 2: Climatic zones of Australia. Arid regions with dry summers in the inside (light orange and light yellow) and wet areas at the coast (dark orange, light and dark blue). Taken from [5]

tively. Furthermore, soil attributes were taken from the Soil and Landscape Grid of Australia (SLGA) and are sampled at a 90 meter resolution. SLGA contains coarse fragments (`cfg`), soil profile depth (A- and B-horizons) (`des`), clay (`cly`), silt (`slt`) and sand (`snd`) content in percent, as well as the soil organic content (`soc`) also in percent.

As this study focuses exclusively on the effects of subsurface backscattering, the data has been filtered to only include landcover types constituting bare soil or sparsely vegetated areas, to reduce noise from other potentially anomalous effects.

2.1 Additional Data

To further enhance the explanatory power of the analysis and capture a wider range of environmental factors that may influence the subsurface backscatter anomaly, several additional data attributes were added. These include elevation, slope, land surface temperature, and precipitation. Each of these variables provides unique insights into the physical and climatic conditions that may affect soil moisture dynamics and the occurrence of subsurface scattering. In the following, I outline the motivation to include each of these attributes and their relevance to the study.

2.1.1 Elevation and Slope

Elevation is a critical factor influencing soil moisture distribution and therefore potential subsurface backscatter anomalies. Studies have shown that higher elevations are associated with cooler temperatures and longer periods of soil water availability. This is due to the combined effects of cooler climates and reduced evaporation rates at higher elevations, which enhance soil moisture retention. In contrast, lower elevations tend to be warmer and drier, with shorter periods of soil water availability, especially in late spring and autumn. Such variations in soil moisture can affect the dielectric properties of the soil and therefore the backscatter signal. [6]

Slope, on the other hand, affects water runoff and infiltration rates. Steeper slopes may lead to faster runoff, reducing soil moisture content and potentially increasing the likelihood of subsurface scattering due to drier surface conditions. [7]

It is important to note that the elevation and slope data were derived from a Digital Elevation Model (DEM). While the Geometric Terrain Correction (GTC) of the model accounts for geometric distortions in the microwave data, it does not completely eliminate radiometric differences caused by terrain geometry, such as variations in the incidence angle of incoming microwaves. These geometric effects can influence backscatter signals independently of soil dielectric properties, as the incidence angle and surface reflectivity vary with topographic changes. [8] This discrepancy highlights the need for careful interpretation of the relationships between terrain features, soil properties, and backscatter signals. Future studies could focus on methods to better disentangle the effects of terrain geometry from soil dielectric properties on microwave backscatter.

By incorporating elevation and slope data, this study aims to better understand how topographic variations contribute to the occurrence of subsurface backscatter anomalies in arid regions.

2.1.2 Land Surface Temperature

Land surface temperature (LST) is a key variable in the energy and water balance of the Earth's surface. It directly influences soil moisture content through evaporation and transpiration processes. Higher LST can lead to increased evaporation rates, resulting in drier soil surfaces, which are more prone to subsurface scattering. [9] Additionally, LST might serve as an indication for the presence of surface features such as rocks or bare soil, which are known to contribute to the subsurface backscatter anomaly. The inclusion of LST data addresses the thermal dynamics of the soil surface and its effect on the backscatter signal, providing a more complete understanding of the drivers of the anomaly.

For this study, the mean LST for 2023 from the MODIS[1] dataset was used to capture average thermal conditions. While this provides a useful overview, it does not account for seasonal temperature changes, which could also impact the anomaly. A more detailed temporal analysis of LST could offer deeper insights but was beyond the scope of this project. Future research could explore LST data over time to better understand how temperature variations influence subsurface scattering, especially in arid regions.

2.1.3 Precipitation

Precipitation is a fundamental driver of soil moisture dynamics and plays a critical role in shaping soil dielectric properties. In arid regions, where subsurface backscatter anomalies are most prevalent, precipitation patterns are often characterised by infrequent and irregular events. These sporadic rainfall episodes, with long dry periods in between, can lead to significant fluctuations in soil moisture. In particular, long dry spells contribute to the formation of hard surfaces or distinct soil horizons beneath the topsoil, which can be favourable for subsurface scattering as microwaves penetrate the dry surface layer and reflect from these underlying features. [1]

For this study, the mean precipitation for the year 2023 from the CHIRPS[2] dataset was used as a proxy for the climatic conditions which have an influence on the variability of soil moisture. While a more detailed temporal analysis, particularly focusing on irregular precipitation events and extended dry periods, is expected to provide deeper insights into the drivers of subsurface backscatter anomalies, such an approach was beyond the scope of this project. Instead, mean precipitation serves as a representative measure of overall water availability.

Future research projects could benefit from incorporating high resolution temporal precipitation data to investigate the impact of specific rainfall patterns and dry spells on soil moisture dynamics and subsurface scattering. This would provide a more nuanced understanding of how short-term climatic variations influence the occurrence of these anomalies, particularly in arid and semi-arid regions.

[1] https://modis.gsfc.nasa.gov/data/
[2] https://www.chc.ucsb.edu/data/chirps

3 Data Preprocessing

The preprocessing phase was crucial to ensure the quality and suitability of the data for the subsequent analysis. Below, I outline the key steps taken during this phase:

3.1 Correlation Analysis

Correlation analysis was used to identify relationships between variables. Not surprisingly, high correlations were observed between soil moisture attributes (`sm`, `sm005`, `sm095`) and also between soil composition attributes (`cly`, `slt`, `snd`). These correlations are expected, as the soil composition attributes directly influence each other. Apart from this, a positive correlation was also found for `precipitation` and `soc`, as well as for `LST` and `cfg`, which could be useful for defining effective interactions during the modeling phase. On the other hand, the remaining variables did not show significant correlations, indicating their independence and potential to add explanatory power to the model.

3.2 Filtering and Encoding

To focus on the topsoil layer, which is most relevant for subsurface backscattering, the `soil_depth` attribute was filtered to include only the topsoil layer (0–5 cm). In addition, the coarse fragment content (`cfg`) was analysed, showing that the fifth category (50-90%) accounted for less than 1% of the observations. Given its negligible representation, only the first four categories of `cfg`, very few (0-2%), few (2-10%), common (10-20%) and many (20-50%), were retained for analysis, while the fifth category was merged with the fourth to serve as a category defined by 'more than 20%'. These categories were one-hot encoded, creating a separate binary column for each category to facilitate their use in modeling.

3.3 Outlier Detection and Removal

Outliers were identified and addressed to ensure the robustness of the dataset. In particular, a small number of extremely negative values in the `elevation` column were identified and removed. These values were considered unrealistic for the study area and were likely due to errors in data collection or processing. As only a few observations were affected, the exact reasons for these outliers were not investigated further. Their removal was necessary to prevent skewing the analysis.

3.4 Scaling and Feature Engineering

To standardize the data and ensure that all variables were on a comparable scale, the dataset was scaled using `StandardScaler`[3]. This step is particularly important for models that are

[3]https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html

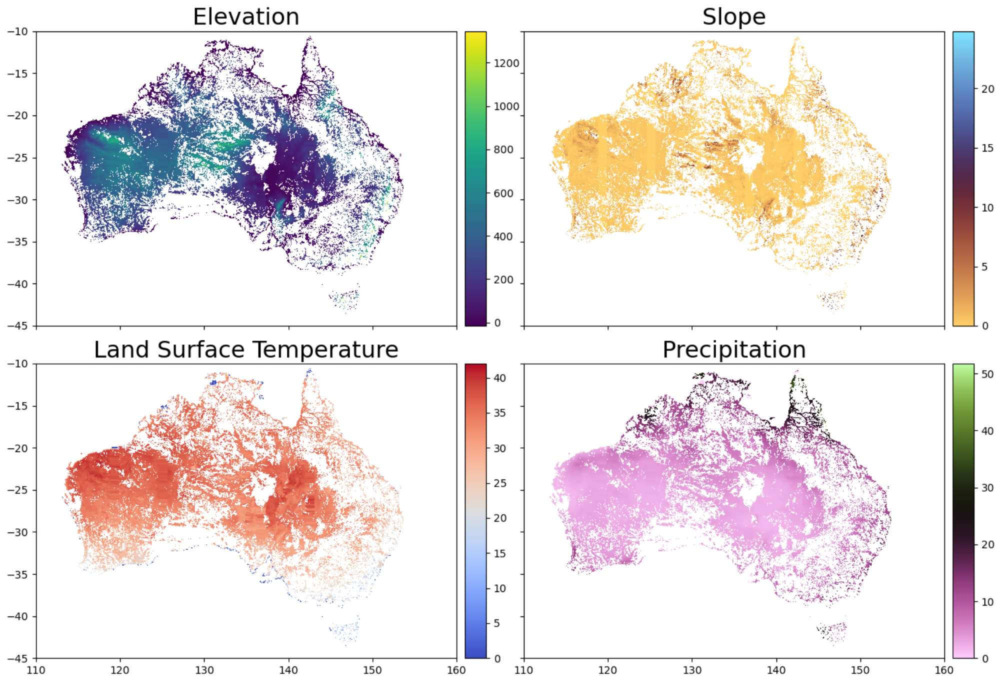

Figure 3: Visualization of the additionally collected data.

sensitive to the magnitude of input features, such as linear regression. Additionally, interaction terms were created using `PolynomialFeatures`[4] with `interaction_only=True`. This approach allowed us to capture potential interactions between variables without introducing higher-order polynomial terms, which could lead to overfitting and add noise.

By following these preprocessing steps, I ensured that the dataset was clean, standardized, and ready for subsequent analysis, enabling a more accurate investigation of the subsurface backscattering anomaly.

3.5 Visual Exploration of the Data

After preprocessing, all data were visually explored to gain a better understanding and to ensure plausibility. This step was particularly important for the additional data collected, as these variables were not part of the original dataset. Visual inspection helped to confirm that the data had been processed correctly and that there were no obvious errors or inconsistencies.

For example, elevation was consistent with expectations for the study area, noting the mountainous regions of Australia[5], while land surface temperature and precipitation values reflected the climatic conditions of arid and semi-arid regions compared to humid coastlines. Figure 3 summarizes these visualizations, providing an overview of the characteristics of the

[4]https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.PolynomialFeatures.html
[5]https://earthobservatory.nasa.gov/images/5100/australia-shaded-relief-and-colored-height

data and supporting the validity of the preprocessing steps.

4 Modeling

A systematic modeling approach was used to address the research question of investigating rock fragments and distinct horizons as drivers of the subsurface backscatter anomaly. This section outlines the methodology, model selection and evaluation process, and the exploration of non-linear models to capture complex relationships in the data.

The modeling strategy followed a top-down approach, starting with the maximum number of attributes and their interactions and gradually refining the model by removing variables with less predictive power. A modeling pipeline was implemented to train and evaluate models using cross-validation. Performance metrics such as Root Mean Squared Error ($RMSE$) and R^2 were reported. Feature importance was assessed using average model coefficients for linear models, impurity-based importance scores for non-linear models and the model-agnostic LIME explainability method, which approximates single predictions of any machine learning models with local linear ones, thus allowing local interpretation.

4.1 Linear Models

The analysis began with linear models, including Multiple Linear Regression, Partial Least Squares (PLS), Ridge Regression, Elastic Net, and Lasso Regression. These models were chosen for their high interpretability, which is crucial for understanding the underlying physical processes driving the subsurface backscatter anomaly. However, as highlighted by Wagner et al. [3], the highly correlated nature of environmental processes posses a significant challenge for modeling. To address this, regularization techniques (Ridge, Lasso, and Elastic Net) and PLS regression were employed to handle multicollinearity and improve model stability.

The modeling approach followed a top-down strategy, starting with all available variables and their interaction terms. Initial models were trained without interaction terms to establish a baseline. Subsequently, the variables and interaction terms with the lowest predictive power were progressively eliminated based on the regularization techniques. This iterative process began with a maximum of 120 features, including all possible two-way interactions and the 15 variables themselves. After the first model fitting process the elastic net model with an L1 ratio of 0.6 had only terms containing a `cfg` variable in the top 6 features. In total, the 20 combinations with the highest absolute model coefficients were used for the next round. This was progressively refined until only 8 input features remained for the final linear models, all of which improved in terms of RMSE and R2 during the selection process.

In addition, a separate analysis was carried out to determine the optimal number of components for the PLS regression, which at this stage proved to be one of the best performing linear models. The results in Figure 4 indicate that the model improves as the number of components increases, as expected, since each additional component captures more of the covariance between the explanatory variables and the target. However, it can be seen that the performance plateaus for $n > 5$, suggesting that additional components contribute

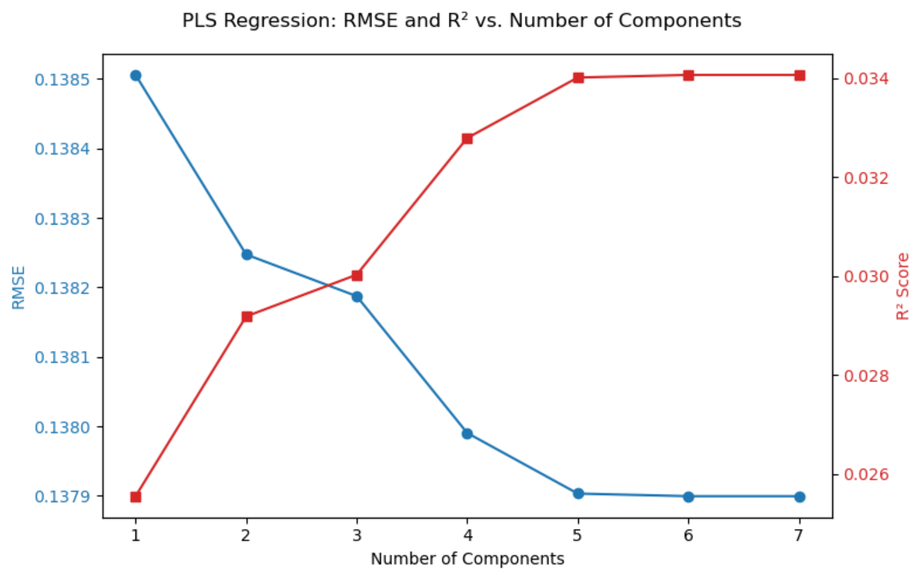

Figure 4: Results of the analysis to optimize number of components for PLS regression. Performance plateaus for $n > 5$.

minimal predictive power, but make the model more complex and may lead to overfitting. Therefore the optimal number of components was determined to be 5 for the final set of 8 variables based on the cross-validation RMSE and R2 scores. Despite this optimization, the predictive power of the linear models remained relatively low, with all models showing comparable performance in terms of $RMSE$ and R^2 values. A more detailed evaluation of these results is provided in Section 4.3.

The final linear models retrieved from the described selection process included the following variables and interaction terms:
`cfg_2`×`cfg_4`, `cfg_2`, `cfg_4`, `slt`×`cfg_4`, `slt`, `LST_Day_1km`×`cfg_4`, `elevation`×`cfg_4` and `sm005`×`cfg_4`.
Notably, 7 of these 8 variables are interaction terms involving a `cfg` category or `cfg_2` and `cfg_4` themselves, underscoring the importance of coarse fragments in explaining anomalous behaviour.

However, the limited predictive power of the models suggests that the relationships may be non-linear or that additional factors not included in the model are influencing the variable of interest. While the linear models provided valuable insights into the relationships between variables and highlighted the importance of coarse fragments as a driver of anomalous behaviour, their poor predictive performance underscores the need for more advanced modeling approaches to capture the complexity of backscatter anomalies. This limitation motivated the exploration of non-linear models, as discussed in the following section.

4.2 Non-Linear Models

Given the limited predictive performance of linear models, non-linear methods were explored to better capture the complex relationships in the data. For this analysis a random forest regressor with 10 trees and a maximum depth of 10, as well as a XGBoost regressor with 20 estimators were used. Similar to the selection process for linear models, an initial baseline model was established without interaction terms. In principle, since both random forests and XGBoost models inherently capture non-linearities, explicit interaction terms are not necessarily required. [10] However, their inclusion can still be beneficial if they provide additional informative features.

For the baseline models, both non-linear methods marginally outperformed the linear models. This trend persisted when using the non-linear methods with the features selected in the linear model feature selection process, though the performance improvement in this scenario was even smaller. Subsequently, an analysis incorporating various interaction terms was conducted. Due to the maximum depth constraint (10) in the random forest model, it could not leverage all variables simultaneously, but should instead prioritize those with the highest predictive power. The primary goal was to compare the features identified as important by linear and non-linear models and assess their consistency with the physical observations described by Morrison et al. [1]

This analysis was conducted using the built-in feature importance attribute of the scikit-learn random forest regressor, which quantifies the reduction in impurity attributable to each feature. Among the most important features, combinations of coarse fragments with elevation, land surface temperature, and precipitation were frequently selected. These three factors, along with `sm005`, were consistently among the highest-ranked in feature importance scores. However, it is crucial to note that, for instance, elevation influences backscattering data not necessarily through subsurface backscattering itself, but rather by altering the incidence angle of the radar signal. Additionally, soil horizons (`des`) and their interactions with `precipitation`, `cfg_2`, and `cfg_4` were among the top-ranked features, appearing at positions 8, 10, and 12, respectively, in the feature importance ranking. These results provide strong evidence that coarse fragments and distinct soil horizons play a crucial role in subsurface backscattering anomalies, aligning well with existing research.

Despite these insights, feature importance alone does not reveal the directionality or precise nature of these relationships. To further investigate these interactions, the advanced explainability method LIME was applied to improve interpretability. LIME (Local Interpretable Model-agnostic Explanations) fits a local linear approximation to explain individual predictions of otherwise complex models. The analysis was conducted multiple times using the features previously identified as relevant in the linear model selection process.

LIME results indicated that:

- A high concentration of coarse fragments (`cfg_4`) combined with silt (`slt`) was inversely related to the predicted value (`rsub`), suggesting a weakening effect on backscattering anomalies.

- A combination of high concentration of coarse fragments (`cfg_4`) with higher land surface temperature LST resulted in higher predicted values.

- The absence of high coarse fragment content (cfg_4) had a negative impact on predictions.

These findings are highly consistent with those reported by Morrison et al.(2020), as both analyses identify coarse fragments as a crucial factor influencing the subsurface backscattering anomaly. Additionally, the positive relationship between high land surface temperature and stronger predictions suggests drier soils as an indicator, which Morrison et al. identified as a key driver as well. Furthermore, the strong negative correlation between silt and sand in this dataset (-0.81) implies that an inverse relationship with silt corresponds to a positive relationship with sand, reinforcing Morrison et al.'s conclusion that sandy soils enhance the anomaly. [1]

While this project focuses primarily on linear methods, the exploration of non-linear models provided valuable insights into the impact of coarse fragments and distinct soil horizons on the subsurface backscattering anomaly. By overcoming the limitations of linear models, non-linear approaches demonstrated their potential in capturing complex interactions. However, this increased model complexity of these models comes at the cost of reduced interpretability. Future work could focus on refining non-linear models and using more advanced interpretability techniques to bridge this gap, enabling a deeper understanding of the drivers behind the subsurface backscattering anomaly.

4.3 Evaluation

To assess the performance of the regression models, standard evaluation metrics, namely Root Mean Squared Error ($RMSE$) and Coefficient of Determination (R^2) were used. $RMSE$ provides insight into the magnitude of prediction errors in the original unit of measurement, while R^2 quantifies the proportion of variance in the target variable explained by the model. In addition to these metrics, the model coefficients across different models were compared to identify key drivers of the subsurface backscatter anomaly.

The best performing linear models, Partial Least Squares (PLS) regression, achieved an $RMSE$ of approximately 0.138 and an R^2 value of 0.034. As shown on the right hand side of Figure 5, the model's predictions have low variance and are clustered tightly around the mean, indicating a tendency towards underfitting. This behavior was consistent across all linear models regardless of the input features. Notably, several experiments with linear models yielded negative R^2 values, suggesting that the models performed worse than a naive baseline (i.e., a horizontal line representing the mean of the target variable).

The left panel of Figure 5 summarizes the coefficients of the final PLS model, where the values reflect the contribution of each feature after accounting for five latent components. The coefficients indicate the strength and direction of the relationship between the input features and target variable. Since all input features were standardized before model fitting, the coefficients are directly comparable. Negative coefficients, such as those for cfg_2×cfg_4 or slt, imply that an increase in these variables correspond to a decrease in Rsub, resulting

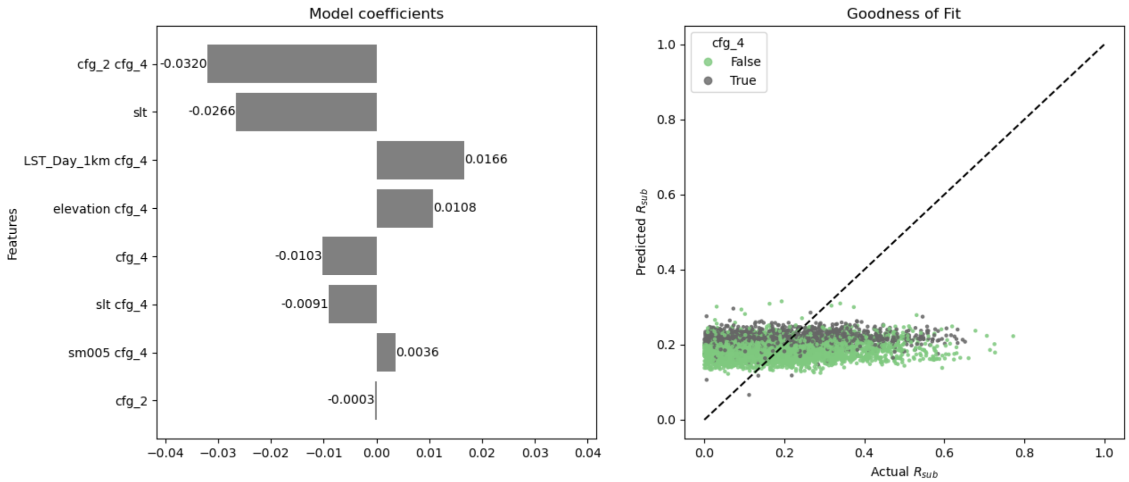

Figure 5: PLS model coefficients and predicted values versus actual values of the target variable `Rsub`

in an inverse relationship. Similar to the results of the LIME explainability method for non-linear models in section 4.2, we see an inverse relationship between silt and the target variable. Note that this suggests a positive relationship for sand, which has a strong negative correlation with silt. Furthermore, the positive coefficient for land surface temperature supports the hypothesis that higher temperatures, which often correspond to drier soils, contribute to subsurface backscatter anomalies. The role of coarse fragments (`cfg`) is less straightforward due to the presence of interaction terms, compounded by class imbalance. However, an examination of prediction errors (Figure 6) indicates that the model performed particularly well in regions with a high concentration of coarse fragments, such as western Australia. Nevertheless, a positive relationship would still have been expected. Finally, elevation was also found to have a positive relationship with the target variable, but as discussed earlier, this is not necessarily due to subsurface scattering alone.

In summary, linear models not only repeatedly highlighted the importance of interaction terms involving coarse fragments (`cfg`), which is consistent with findings from existing research [3], but also agreed on the direction in which the various factors influence the anomalous behaviour. [1] These results strongly suggest that coarse fragments, particularly their interactions with other factors of soil composition and the dryness of the soil, play a significant role in driving the anomaly.

5 Conclusion

The aim of this study was to investigate the drivers of subsurface backscatter anomalies in dry, arid regions where traditional microwave derived soil moisture retrievals show anomalies due to subsurface scattering. To improve the explanatory power of the analysis, additional

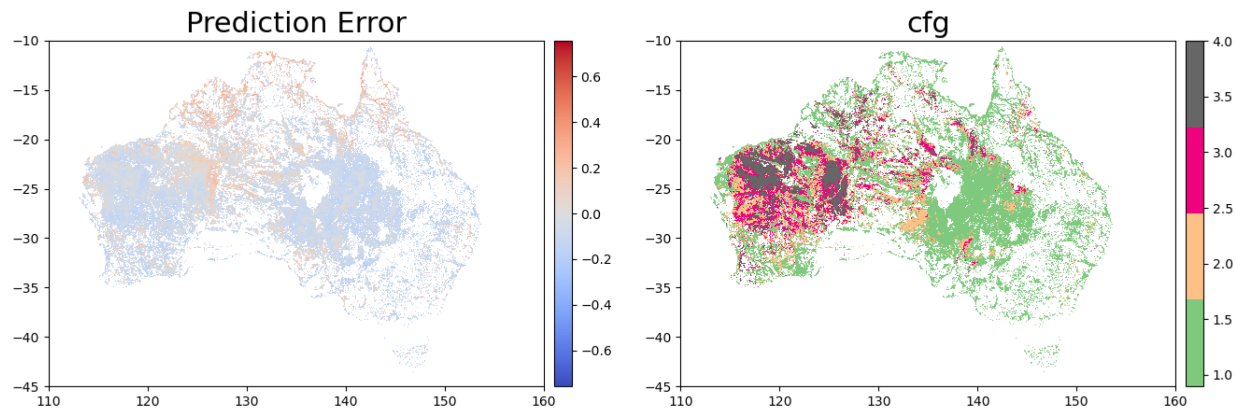

Figure 6: Prediction error compared to the occurrence of coarse fragments (1: very few - 4: many)

environmental data - elevation, slope, land surface temperature and precipitation - were included as these factors influence soil moisture dynamics and thus potentially the conditions under which subsurface scattering occurs.

The results underscore the significant role of coarse fragments and distinct soil horizons in shaping subsurface backscattering behavior. However, these factors alone do not fully determine the presence of subsurface scattering. Interactions with silt content and land surface temperature emerged as strong predictive indicators. Despite these insights, the study also revealed key limitations in both linear and non-linear modeling approaches. While linear models provided interpretability, they struggled to accurately capture the complex, non-linear interactions among environmental variables, leading to suboptimal predictive performance. Non-linear models — particularly Random Forest — demonstrated a slight improvement in predictive accuracy, but their reduced transparency made it challenging to clearly interpret the physical mechanisms underlying the observed anomalies. The results of the feature importance analysis and the LIME explainability methods do support existing research, however. It is noteworthy that this is particularly evident in both modeling approaches.

These findings highlight the need for future research to address three critical areas. First, incorporating high-resolution temporal precipitation data and seasonally varying land surface temperature could better capture the influence of seasonal and climatic variations on subsurface backscattering dynamics. Second, developing models tailored to specific climatic zones may enhance predictive accuracy by accounting for regional environmental differences. Finally, advancing explainability techniques for non-linear models would improve the interpretability of complex interactions, making remote sensing methods for soil moisture estimation more robust and applicable across diverse environments. By addressing these challenges, future studies can significantly refine our understanding of subsurface scattering processes and improve remote sensing techniques for soil moisture retrieval.

References

[1] K. Morrison and W. Wagner. Explaining anomalies in sar and scatterometer soil moisture retrievals from dry soils with subsurface scattering. *IEEE Transactions on Geoscience and Remote Sensing*, 58(3):2190–2197, Mar 2020.

[2] W. Wagner et al. Widespread occurrence of anomalous c-band backscatter signals in arid environments caused by subsurface scattering. *Remote Sensing of Environment*, 276:113025, Jul 2022.

[3] W. Wagner et al. Mapping subsurface scatterers from sar backscatter time series. In *EUSAR 2024*, 2024.

[4] Roland Lindorfer, Wolfgang Wagner, Sebastian Hahn, Hyunglok Kim, Mariette Vreugdenhil, Alexander Gruber, Milan Fischer, and Miroslav Trnka. Global scale maps of subsurface scattering signals impacting ascat soil moisture retrievals, Aug 2023.

[5] Bureau of Meteorology Australian Governemnt. Climate classification maps, 2025. https://http://www.bom.gov.au/climate/maps/averages/climate-classification/ [Accessed: 2025-03-20].

[6] Jeanne C. Chambers Bruce A. Roundy. Effects of elevation and selective disturbance on soil climate and vegetation in big sagebrush communities. *Ecosphere*, 12:e03377, 2021.

[7] I. V. Florinsky. *Digital Terrain Analysis in Soil Science and Geology*. Academic Press, 2016.

[8] Wolfgang Wagner et al. Earth observation data science cookbook, 2025.

[9] Kang Jiang et al. Combined influence of soil moisture and atmospheric humidity on land surface temperature under different climatic background. *iScience*, 26, 2023.

[10] Nidhi Sahai. Random forest regression – how it helps in predictive analytics?, 2023. Accessed: 2025-03-12.