SUPPORTING SEISMIC DATA SURVEY DESIGN THROUGH THE INTEGRATION OF SATELLITE-BASED LAND COVER MAPS

Liang Tian[1], *Naveed Akram*[2], *Rocco Sedona*[3], *Nikos Savva*[2,5],
Morris Riedel[1,3], *Gabriele Cavallaro*[1,3], *Eric Verschuur*[4]

[1]School of Engineering and Natural Sciences, University of Iceland, 107 Reykjavik, Iceland
[2]The Cyprus Institute, 2121 Aglantzia, Cyprus
[3]Jülich Supercomputing Centre, Forschungszentrum Jülich, 52428 Jülich, Germany
[4]Delft University of Technology, Faculty CEG, 2628 CN Delft, The Netherlands
[5]University of Cyprus, Aglantzia, Nicosia, 2109, Cyprus

ABSTRACT

Seismic Imaging (SI) survey design for onshore applications faces challenges such as accessibility and poor data quality due to unexpected (near-)surface conditions. In this paper, we explore the correlation between the surface conditions provided by Land-Cover (LC) maps generated using Remote Sensing (RS) data and different settings of Seismic Processing (SP) parameters. The study involves a 2D seismic line related to geothermal exploration from the Netherlands.

***Index Terms*—** Remote Sensing, Seismic Data Processing, Seismic Imaging, Deep Learning, Supercomputing

1. INTRODUCTION

SI is the process of creating an image of the Earth's subsurface from seismic data. While SI is a powerful tool for discovering the characteristics of subsurface geology, it has some challenges, especially for the on-shore situations. When we design the seismic survey, safety, and accessibility are always an important topic [1]. Moreover, different surface and near-surface conditions may affect the quality of the acquired seismic data [2]. Therefore, the survey design incorporates choices based on expected seismic data quality.

A common method of assessing surface conditions for seismic surveys is physical site visits and ground surveys, where geologists inspect soil types and identify potential obstacles. However, these are costly and time-consuming. RS can help to solve this problem. Satellite imagery offers high-resolution images of the Earth's surface including dynamic environmental changes, land cover maps, terrain, and topography. Many applications use satellite data in seismology, such as mapping active seismic faults on satellite images to monitor and predict forthcoming earthquakes based on rupture dynamics of these fault zones [3], interpretation of tectonic and stratigraphic zones [4], imaging landslide structures in mountain areas including ground motion displacement [5]. Contemporary Earth Observation (EO) programs like European Space Agency's Copernicus offer open and freely accessible high-resolution, multi-temporal, and multi-spectral RS data globally. This satellite data can facilitate an initial assessment of accessibility and geological features, which is advantageous for Seismic Acquisition Design (SAD). Integrating the information acquired from seismic and satellite data to improve seismic processing workflows involves a multidisciplinary approach that combines RS, geophysics, and geological studies. Laake et al. [6] already showed the relation between satellite images and the estimated time shifts (so-called 'statics') to correct for low, near-surface seismic velocities.

In this work, we explore the correlation between surface conditions, seismic processing parameters, and observed noise content based on a 2D seismic line from the Netherlands related to geothermal exploration. Predicting seismic attributes from satellite data could enhance the accuracy and reliability of subsurface models, providing a better understanding of Earth's subsurface structures and their corresponding physical properties. This can also support the design of seismic surveys by making seismic data collection more effective and efficient, which is important for low-budget seismic applications related to the energy transition.

2. METHODOLOGY

2.1. Remote Sensing

The surface conditions are studied with RS by generating a LC map using a modified version of the classification system proposed by Paris *et al.* [7]. The system includes two main steps. First, the time series of multi-spectral satellite data are collected from the area where the seismic data are acquired. After the clouds are removed from the metadata, a reliable labeled training set is extracted from the RS acquisitions, which is then used to train Machine Learning (ML) or Deep Learning (DL) models. In this work, we use a Transformer DL model [8] to predict the LC map. The workflow can scale on High-Performance Computing (HPC) systems, as it relies on parallel algorithms, such as those proposed by Tian *et al.* [9].

2.2. Seismic Processing and Imaging

For onshore active SI applications, a major role is attributed to the pre-processing of the seismic measurements. Two main factors influence the expected image quality: the presence of deterministic noise in terms of low-velocity surface waves with strong amplitudes. This noise is largely determined by the so-called weathering layer, where the speed of sound of such surface waves is strongly coupled to the soil type of this upper layer [2]. The location and spatial sampling of seismic receivers largely determine the ability to remove these noise events and reveal the desired subsurface reflections. Next, this low-velocity weathering layer also determines small time delays observed in the seismic data, known as statics, that need to be estimated and removed before further imaging steps can be applied [10]. Therefore, the SAD for onshore applications is guided by two basic principles [11]: the ability to suppress noise that is mostly generated in the near-surface region and the ability to illuminate the target area. This information significantly facilitates the SAD process.

2.3. Integration

In this paper, we propose to integrate the LC map information with SI, especially in the design phase, as follows: First, the intended area is mapped using satellite image data and is converted to LC maps using a DL approach, trained on previous data from the same or similar areas. Next, the important seismic data properties (such as expected surface-wave velocities or noise content) can be mapped according to previous correlations made from observed seismic data and LC maps. Finally, based on these correlations, optimized decisions can be made on the SAD regarding the location and sampling of the seismic sources and/or receivers.

The work expands its significance to a variety of applications in geo-energy and geotechnical industries, such as finding subsurface caverns for CO2 or green hydrogen storage, subsurface characterization and monitoring, and shallow surface investigation for civil engineering applications.

3. EXPERIMENTS AND RESULTS

The work is subdivided into three parts: seismic field data analysis, the LC map generation, and the study of the correlation between these two types of data.

3.1. 2D seismic line for geothermal exploration

Because of the urgency for the energy transition, 2D seismic lines have been acquired across the Netherlands in the so-called SCAN program [12] to identify locations for geothermal energy generation using deep subsurface formations.

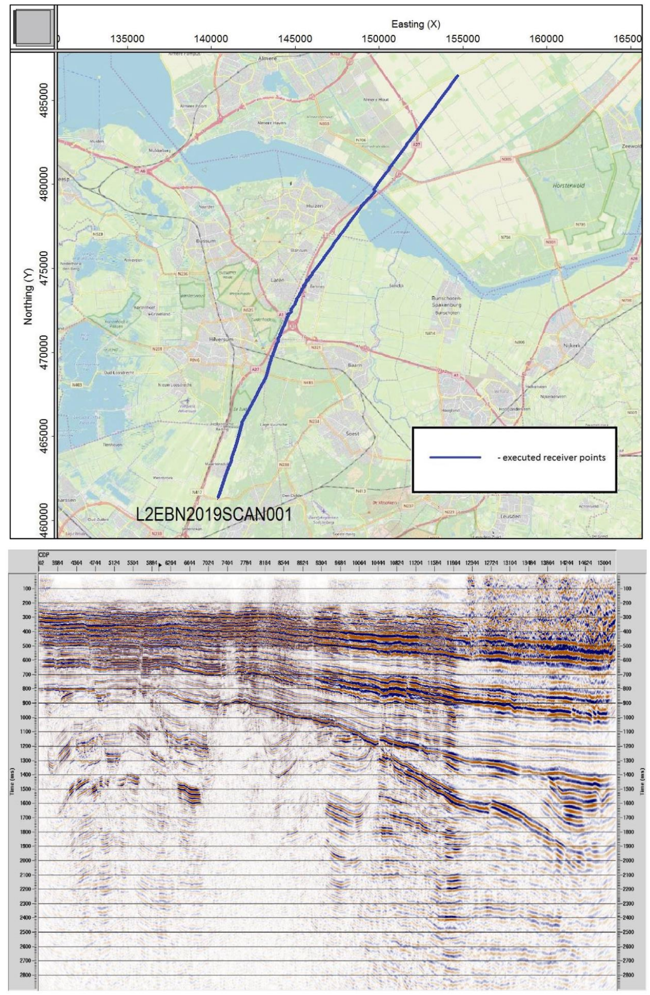

Fig. 1: Zoom of the map of the Netherlands with the location of the seismic line indicated in blue and a processed seismic image along this line (starting at the lower left) displayed below. Both figures are taken from [13]

An important study area in the Netherlands is selected both for seismic and satellite imagery considering its potential for geothermal reservoirs. The data from a twenty-nine-kilometer-long two-dimensional seismic line, originally acquired for a geothermal reservoir survey, is used for the proposed analysis. The map with the line and the corresponding seismic section are shown in Figure 1.

We extracted the following information from the raw data (i.e., the complete seismic gather for each shot number): A map measuring the local speed of sound of the surface waves (V_s) as a function of receiver station (rcv nr) and source or shot number. This is visualized in Figure 2 as the left-most panel. We considered the 200 closest receiver channels for each shot, yielding this band-matrix structure. We can recog-

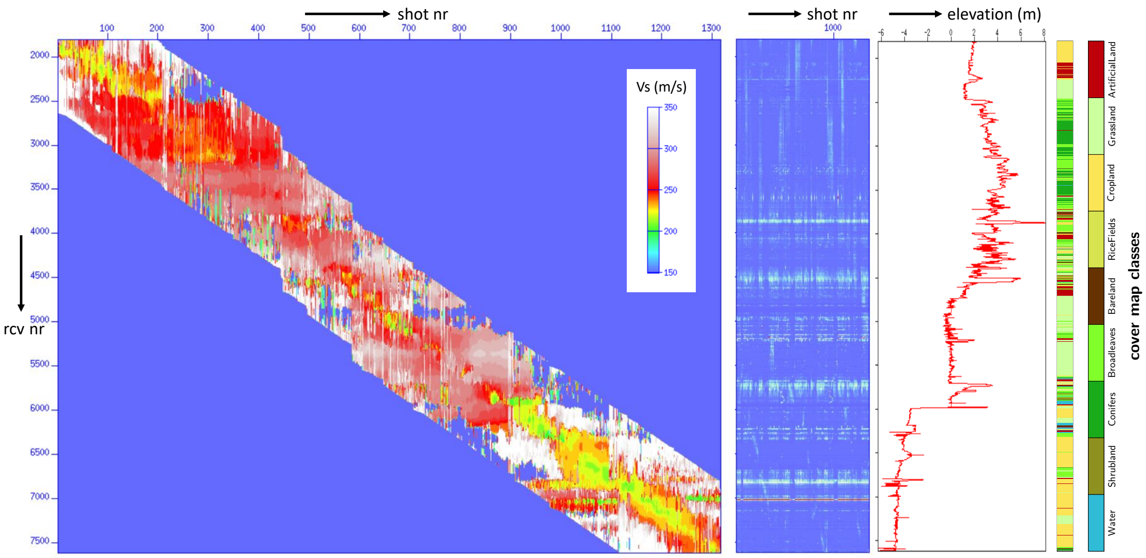

Fig. 2: Seismic and satellite image attributes used in the analysis, with "rcv nr" being the position/station location along the line. (left) estimated V_s as a function of shot and rcv nr., (left middle) noise as a function of shot and rcv nr., (right middle) receiver elevation, and (right) Land-Cover Map Class (LCMC) along the line with its color map.

nize different regions where V_s is low (yellow color) or relatively high (red-white color). Similarly, we calculated the noise content as a function of the shot and receiver number (left-middle panel). We also extracted the elevation for each receiver station (right-middle panel). We identified 22 regions along the line, and for each region we collected the lowest and highest V_s value and the noise level (mapped on a scale from 0 to 10).

3.2. Satellite imaging-based land cover maps

For investigating the surface, satellite data that are collected from March 2018 to March 2019 are utilized to produce the LC maps for this region. The classification code employs a PyTorch-based implementation of the Transformer model [14] for DL. This model's training leverages the PyTorch Distributed Data Parallel framework to efficiently scale across multiple GPUs of JURECA-DC[1]. With a 10-meter spatial resolution, the surface conditions are represented through 9 classes (see Figure 2, right). The LC maps have achieved a local accuracy of 78.99.

3.3. Classification using a Random Forest approach

From satellite image data we can extract two types of information: the elevation values at the seismic station locations and the LCMC values that have been determined. Both are expected to correlate with the SP parameters, which could mean that for e.g. acquisition design purposes, the SP parameters can be predicted from the satellite data. We investigate this via a Random Forest model [15] and Figure 3 shows the so-called confusion matrix for the prediction of the noise level (ranging from 0-10) at each location based on the LCMC, the elevation, or both. The confusion matrix indicates how well each true noise label is determined from the input variables (being LCMC and/or elevation). We observe a significant improvement in the accuracy value for the case of the combined information, as given in Table 1. As a result, the confusion matrix gets more of a diagonal structure (Figure 3, bottom).

[1]JURECA-DC: `https://www.fz-juelich.de/en/ias/jsc/systems/supercomputers/jureca`

Table 1: Accuracies of the Random Forest prediction of three seismic parameters (Noise level, lowest V_s, highest V_s) from three types of input information.

Input	$Noise$	$V_{s,low}$	$V_{s,high}$
Classes	0.50	0.46	0.36
Elevations	0.71	0.64	0.59
Classes & Elevations	**0.79**	**0.76**	**0.72**

We did a similar analysis for the parameters $V_{s,low}$ and $V_{s,high}$. The accuracy values are also shown in Table 1.

4. CONCLUSION

Based on results from the seismic line in the Netherlands, satellite image information can be used to predict SP parameters automatically. The satellite data information includes elevation and the LC map, which is extracted via a classification system. Combining these two types of information yields the highest accuracy for SP parameters, ranging from 0.72 to 0.79. Such accurate prediction of seismic parameters will be valuable in SAD processing, where the locations and density of the seismic stations are crucial parameters.

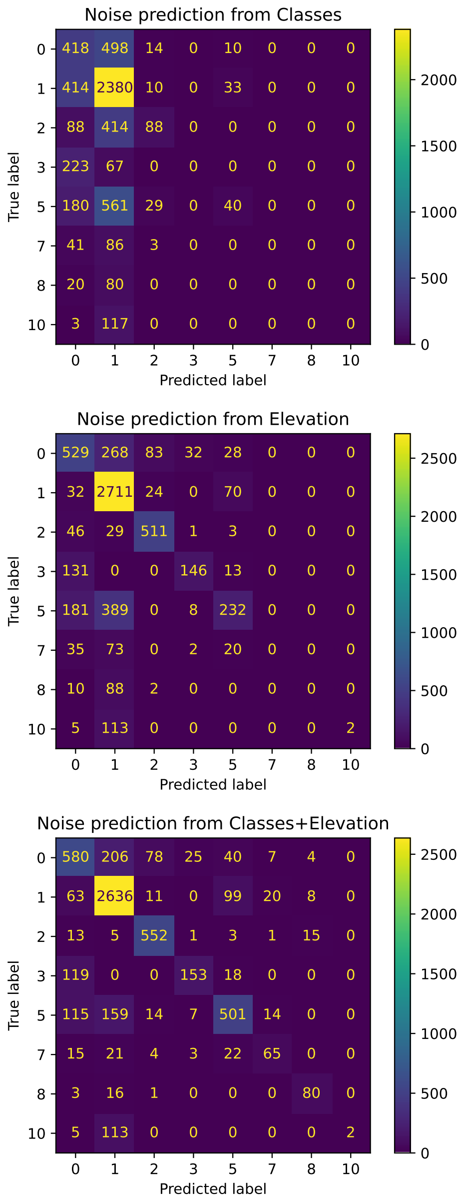

Fig. 3: Confusion matrix for predicted noise content from (top) LCMC information only, (middle) elevation information only, (bottom) LCMC and elevation combined.

5. ACKNOWLEDGEMENTS

This work was performed in the Center of Excellence (CoE) Research on AI- and Simulation-Based Engineering at Exascale (RAISE) receiving funding from EU's Horizon 2020 Research and Innovation Framework Programme H2020-INFRAEDI-2019-1 under grant agreement no. 951733. This work is co-financed by the EUROCC2 project funded by the European High-Performance Computing Joint Undertaking (JU) and EU/EEA states under grant agreement No 101101903. We thank Dr. Jasper Verschuur, Oxford University, for his help with the Random Forest calculations. The SCAN lines processing reports and sections can be found at: `https://www.nlog.nl/scan-2d-seismische-data`.

6. REFERENCES

[1] A. Laake, "Applications of satellite imagery to seismic survey design," *Geophysics*, vol. 23, 10 2004.

[2] T. H. Keho and P. G. Kelamis, "Focus on land seismic technology: The near-surface challenge," *The Leading Edge*, vol. 31, no. 1, pp. 62–68, 2012.

[3] A. Tronin, "Satellite remote sensing in seismology. a review," *Remote Sensing*, vol. 2, 12 2009.

[4] R. Thannoun, S. Fanoosh, and H. Adeeb, "Integration of satellite data processing with seismic sections for tectonic interpretation and modeling for the breaks and omissions of continuous stratigraphic units," vol. 12, 01 2021.

[5] M. Wróbel, I. Stan-Kłeczek, A. Marciniak, M. Majdański, S. Kowalczyk, A. Nawrot, and J. Cader, "Integrated geophysical imaging and remote sensing for enhancing geological interpretation of landslides with uncertainty estimation-a case study from cisiec, poland," *Remote Sensing*, vol. 15, pp. 238, 12 2022.

[6] A. Laake, "Integration of satellite imagery, geology and geophysical data," in *Earth and Environmental Sciences*, I. A. Dar and M. A. Dar, Eds., chapter 21. IntechOpen, Rijeka, 2011.

[7] C. Paris, L. Gasparella, and L. Bruzzone, "A scalable high-performance unsupervised system for producing large-scale hr land cover maps: The italian country case study," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 9146–9159, 2022.

[8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2023.

[9] L. Tian, R. Sedona, A. Mozaffari, E. Kreshpa, C. Paris, M. Riedel, M. G. Schultz, and G. Cavallaro, "End-to-end process orchestration of earth observation data workflows with apache airflow on high performance computing," in *IEEE IGARSS*, 2023, pp. 711–714.

[10] M. Cox, *Static Corrections for Seismic Reflection Surveys*, Society of Exploration Geophysicists, 1999.

[11] G. J. O. Vermeer, *3D Seismic Survey Design, Second edition*, Society of Exploration Geophysicists, 2012.

[12] J. Rehling, "SCAN 2D reprocessing close-out report," Tech. Rep., EBN, 2023.

[13] "Final report GTO-19-C011 EBN 2D test line L2EBN2019SCAN001 data processing," Tech. Rep., EBN, 2019.

[14] M. Rußwurm, C. Pelletier, M. Zollner, S. Lefèvre, and M. Körner, "Breizhcrops: A time series dataset for crop type mapping," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2020.

[15] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.