

Quantifying glacial sediment thicknesses offshore Greenland: Using a neural network to extrapolate spatial data in remote areas

Caroline Brand¹, Judith Elger¹, Katrine J. Andresen^{1,2}, Thomas M. Hansen¹, Victor S. Poulsen¹, Lara F. Pérez², Christoph Böttner³, Paul C. Knutz², Jonas F. Damsgård¹, Vivi K. Pedersen¹

¹ Department of Geosciences, Aarhus University, Denmark; ² GEUS, Geological Survey of Denmark and Greenland, Denmark; ³ Department of Geological Sciences, Stockholm University, Sweden

Glacial erosion rates over a million-year time scale are relatively unconstrained for the Greenland Ice Sheet (e.g. Strunk et al., 2017). Yet these rates can give valuable insights into long-term landscape evolution and glacial dynamics. One approach to constrain the glacial erosion is to quantify glacial deposition volumes (Dowdeswell et al., 2010; Paxman et al., 2019; Pedersen et al., 2021). However, due to Greenland's remoteness and the challenging environmental conditions for seismic acquisition, spatial gaps in the seismic sediment record remain. Consequently, a quantification of glacial sediment volumes is difficult. To address this, we present a first compilation of available and newly interpreted Quaternary marine seismic data with age constraints where scientific boreholes are available. For areas with sparse data coverage, we employ a data-informed artificial neural network to predict Quaternary sediment thicknesses. The neural network is trained using the seismic-derived thicknesses, along with several parameters related to glacial and geomorphological features. This approach allows us to obtain a glacial sediment thickness map, thereby constraining the total volumes of deposition. The predicted sediment thicknesses reveal variations in glacial deposition volumes across different areas of Greenland's continental margin. The steeper southern and parts of the northern Greenlandic continental slope are covered by thinner Quaternary sediments, whereas in west and east Greenland, larger glacial sediment deposits advanced the shelf farther throughout the Quaternary, extending observations of regional seismic studies (e.g., Heirmann et al., 2019; Knutz et al., 2019; Nielsen et al., 2019; Pérez et al., 2018). These differences in glacial deposition highlight the diverse influence of paleoclimatic, oceanographic, and orographic processes on glacial dynamics and the source-to-sink sediment transport. Our results show that the application of machine learning to integrate available seismic records can provide new insights into the temporal and spatial variability of Quaternary glacial erosion and deposition processes in Greenland.

Dowdeswell, J. A. et al. (2010). Rates of sediment delivery from the Fennoscandian Ice Sheet through an ice age. *Geology*, 38(1), 3-6. <https://doi.org/10.1130/g25523.1>

Heirman, K. A., Nielsen, T., & Kuijpers, A. (2019). Impact of Tectonic, Glacial and Contour Current Processes on the Late Cenozoic Sedimentary Development of the Southeast Greenland Margin. *Geosciences*, 9(4), 157. <https://doi.org/10.3390/geosciences9040157>

Knutz, P. C., Newton, A. M., Hopper, J. R., Huuse, M., Gregersen, U., Sheldon, E., & Dybkjær, K. (2019). Eleven phases of Greenland Ice Sheet shelf-edge advance over the past 2.7 million years. *Nature Geoscience*, 12(5), 361-368. <https://doi.org/10.1038/s41561-019-0340-8>

Nielsen, T., Kuijpers, A., & Mørk, F. (2019). Sedimentary processes and seabed morphology of the Southwest Greenland margin. *arktos*, 5(2), 89-104. <https://doi.org/10.1007/s41063-019-00069-5>

Paxman, G. J., Jamieson, S. S., Hochmuth, K., Gohl, K., Bentley, M. J., Leitchenkov, G., & Ferraccioli, F. (2019). Reconstructions of Antarctic topography since the Eocene-Oligocene boundary. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 535, 109346. <https://doi.org/10.1016/j.palaeo.2019.109346>

Pedersen, V. K., Knutsen, Å. R., Pallisgaard-Olesen, G., Andersen, J. L., Moucha, R., & Huismans, R. S. (2021). Widespread glacial erosion on the Scandinavian passive margin. *Geology*, 49(8), 1004-1008. <https://doi.org/10.1130/G48836.1>

Pérez, L. F., Nielsen, T., Knutz, P. C., Kuijpers, A., & Damm, V. (2018). Large-scale evolution of the central-east Greenland margin: New insights to the North Atlantic glaciation history. *Global and Planetary Change*, 163, 141-157. <https://doi.org/10.1016/j.gloplacha.2017.12.010>

Strunk, A., Knudsen, M. F., Egholm, D. L., Jansen, J. D., Levy, L. B., Jacobsen, B. H., & Larsen, N. K. (2017). One million years of glaciation and denudation history in west Greenland. *Nature Communications*, 8(1), 14199. <https://doi.org/10.1038/ncomms14199>

Transparent Framework for Stakeholder-Aware Geothermal Site Suitability Screening in the Netherlands

Karian Bruyndonckx^{1,2}, Hamed Ali Diab-Montero², Eva De Boever², Derek Karssenberg¹, Vera Pijnenburg³, Madlene Nussbaum¹

1 Utrecht University, Utrecht, The Netherlands, 2 TNO – Geological Survey of the Netherlands, Utrecht, The Netherlands, 3 EBN, Utrecht, The Netherlands

Keywords: *geothermal energy, suitability screening, multi-criteria decision analysis, decision support, stakeholder*

In the Netherlands, geothermal energy is expected to play a substantial role in the heat transition, yet deployment remains ~7.5 PJ in 2024, far below the 80 PJ ambition for 2050. A national exploratory study commissioned under the Bodem en Ondergrond covenant shows that barriers already emerge before feasibility studies, when municipalities, provinces and other actors must judge whether development is plausible. Early-stage screening is therefore a multi-actor, multi-dimensional challenge that must align subsurface potential with surface realities. Although subsurface potential is publicly available (e.g., ThermoGIS), screening in practice is still largely technical: socio-economic, regulatory, environmental, built-environment and infrastructure constraints are often addressed late, frequently only downstream in processes that precede or inform programmes such as EBN's SCAN programme. This limits early spatial prioritisation, realistic expectation-setting and stakeholder communication.

To address this gap, we develop and evaluate a transparent, stakeholder-aware spatial screening framework for deep geothermal suitability in the Netherlands that integrates technical, regulatory, environmental and socio-economic considerations in an interpretable GIS-based Multi-Criteria Decision Analysis (MCDA) structure [Malczewski 2006, Raos et al 2022], while making differences between stakeholder perspectives explicit. Using a Design Science Research (DSR) approach with TNO and EBN, criteria are translated into spatial indicators and combined through governed rules; a medallion architecture links source data to harmonised layers and model-ready indicators, and an invariant core stabilises screening semantics so stakeholder perspectives can be explored through bounded configurations without changing the model [Hevner et al 2024]. Evaluation is embedded through expert walkthroughs, sensitivity analysis and policy case studies in Zuid-Holland (mature context) and Noord-Brabant (exploratory), using interpretability, credibility, robustness, decision relevance and traceability as explicit quality dimensions.

The framework shifts screening beyond a single technical suitability surface by distinguishing exclusion, limitation and relative opportunity, clarifying where favourable subsurface conditions are unlikely to translate into plausible development contexts. It makes stakeholder prioritisation logics comparable within one shared structure and produces policy-facing outputs that identify where geothermal development may be promising, and which surface-level bottlenecks dominate at the pre-feasibility stage. Overall, the work underlines that the Dutch geothermal challenge is not only to identify technically promising subsurface conditions, but to make them interpretable considering governance, spatial and socio-economic realities. By formalising these interactions within one GIS–MCDA screening structure, the approach improves early-stage decision-making and communication by reinterpreting geothermal resource information as resource opportunity, dominant limitations and stakeholder-dependent outcomes.

Hevner, A.R., Parsons, J., Brendel, A.B., Lukyanenko, R., Tiefenbeck, V., Chiarini Tremblay, M., and vom Brocke, J. (2024). Transparency in design science research. *Decision Support Systems*, 182, 114236. <https://doi.org/10.1016/j.dss.2024.114236>

Malczewski, J. (2006). GIS-based multicriteria decision analysis: a survey of the literature. *International Journal of Geographical Information Science*, 20(7), 703–726. <https://doi.org/10.1080/13658810600661508>

Raos, S., Hranić, J., Rajšl, I., and Bär, K. (2022). An extended methodology for multi-criteria decision-making process focused on enhanced geothermal systems. *Energy Conversion and Management*, 258, 115253. <https://doi.org/10.1016/j.enconman.2022.115253>

Geo-ASCII: A Geoscience AI-assisted Seismic Conditioning and Iterative Interpretation workflow

Stefan Carpentier¹, Camille Chapeland¹, Suzanne Atkins¹, Boris Boullenger¹, Jan Diederik van Wees^{1,2}

¹ TNO – Geological Survey of the Netherlands, Utrecht, The Netherlands; ² Utrecht University, Utrecht, The Netherlands

Conventional seismic interpretation and fault detection use manual picking, tracked picking, and seismic attributes for horizons and faults using routines such as OpendText auto-tracking and Petrel ant-tracking. This is cumbersome and leads to underutilization of data, as only a limited portion of all data is interpreted/picked in 3D surveys. We propose scaling up the interpretation of seismic data in the Dutch subsurface by interpreting and picking horizons and faults in the full range of data in Dutch 3D surveys instead of just a subset of inlines and crosslines using Machine Learning (ML) algorithms.

Next-generation Machine Learning tools that train and apply neural networks for automated horizon and fault detection have been developed already by market parties. Although generally successful and orders of magnitude faster, there are still many inaccuracies in these ML predictions compared to human interpretation/picking.

We have developed a ready-to-use AI software tool that uses seismic data and logs as input and generates output on predicted seismic horizons, geological units, and faults. For reducing accuracies and improving customization, we have inspected, conditioned, trained, retrained, and corrected state-of-the-art ML algorithms with the expertise and knowledge of human interpreters and available horizons and faults, such as Dutch Deep/Shallow NLOG datasets. We have cast this into a workflow for simple, user-friendly, low-threshold interaction with the ML tools. The application of the ML tools has been applied on onshore shallow to deep reservoirs/aquifers for HTO and geothermal exploration.

The first component of the workflow, AI data conditioning, performs automatic conditioning of seismic data on large amounts of 2D and 3D seismic data by doing denoising, interpolation and broadbanding. For that we use Self-supervised learning (“Noise2Void”, Birnie et al., 2021) and a Generative Adversarial Network (“MDA-GAN”, Dou et al., 2022, (a)).

Secondly, we automatically interpret geological horizons on large amounts of 2D and 3D seismic data. There we employ image segmentation using semi-supervised learning for classifying over- and underburden of given horizons, training on cleaned 3D Inlines and Crosslines (CONSS, Li et al., 2023).

Finally, we automatically interpret faults on large amounts of 2D and 3D seismic data. This is done by image segmentation with semi-supervised learning networks “FaultSSL” and FaultNet (Dou et al., 2022 (b), Dou et al., 2023) for fault detection in seismic data, training on Dutch 3D seismic cubes.

This Geo-ASCII workflow is demonstrated on several seismic datasets for the purpose of shallow and deep geothermal exploration. The workflow of these AI networks is mature and performs well on synthetic and field seismic data. Training on synthetic data always works well, training on field data is variable, the labels are not always consistent or correct and are scale variant. The 3D consistency of the AI predicted geological units, horizons, and faults is very convincing and usable, even in the shallow domain of the Dutch Maassluis and Oosterhout Formations.

Birnie, C., Ravasi, M., Liu, S., Alkhalifah, T., (2021), The potential of self-supervised networks for random noise suppression in seismic data, *Artificial Intelligence in Geosciences* 2, p 47-59.

Dou, Y., Li, K., Duan, H., Li, T., Dong, L., Huang, Z., (2022, (a)), MDA GAN: Adversarial-Learning-based 3-D Seismic Data Interpolation and Reconstruction for Complex Missing, *arXiv:2204.03197*.

Dou, Y., Li, K., Zhu, J., Li, T., Tan, S., Huang, Z., (2022, (b)), MD loss: Efficient training of 3-D seismic fault segmentation network under sparse labels by weakening anomaly annotation, *IEEE Transactions on Geoscience and Remote Sensing*, Volume 60, pages 1–14.

Dou, Y., Li, K., Dong, M., Xiao, Y., (2023), FaultSSL: Seismic Fault Detection via Semi-supervised learning, *Geophysics*, volume 89, Number 3, Pages 1–43.

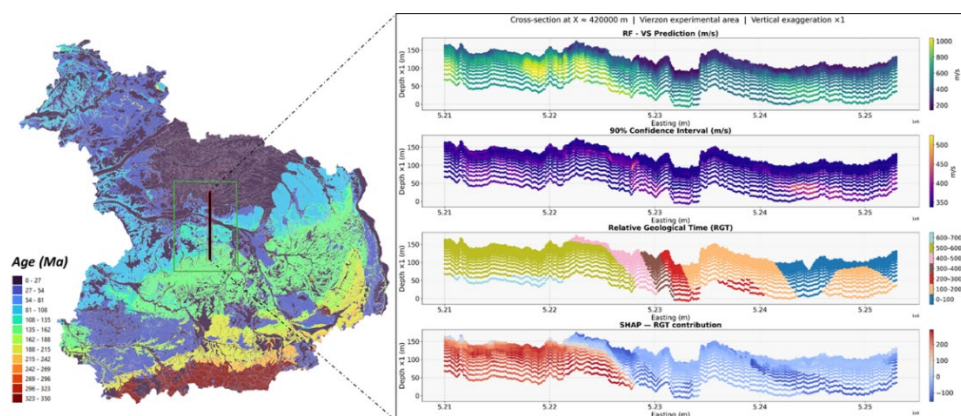
Li, K., Liu, W., Dou, Y., Xu, Z., Duan, H., Jing, R., (2023), CONSS: Contrastive Learning Approach for Semi-Supervised Seismic Facies Classification, *ArXiv 2210.04776v3*.

Uncertainty-Aware 3D Prediction of S-wave and P-wave Velocities from Multimodal Data

Roberto Cilli¹, Guillaume Caumon^{1,2}, Behshad Koohbor¹, Marianne Clausel³, Jihad Bourassi⁴, Pierre-Alexandre Reninger⁴, Aurélie Peyrefitte⁴

1 Université de Lorraine, CNRS, GeoRessources, Nancy. 2 Institut Universitaire de France. 3 Université de Lorraine, CNRS, CRAN, Nancy. 4 Bureau de Recherches Géologiques et Minières (BRGM), Orléans

Accurate characterization of the shallow subsurface is a prerequisite for natural hazard assessment, hydrogeological studies and land use planning. We present a multimodal machine learning framework for the 3D prediction of seismic S-wave and P-wave velocities in an experimental area of 500 km² in the southern part of Paris basin. The framework integrates four heterogeneous data sources: vertical resistivity profiles obtained from airborne electromagnetic data (SkyTEM [1]), interpolated using layer-by-layer 2D Ordinary Kriging; surface lithological descriptions from the BRGM converted into semantic embeddings via a pre-trained SentenceBERT model [2]; a 3D structural geological model built from nearly 10,000 stratigraphic contacts from ~600 borehole using LoopStructural [3], providing relative geological time (RGT); and Earth Observation embeddings (AlphaEarth) [4]. Sparse S-wave and P-wave velocity measurements from boreholes serve as ground truth. A key finding is that the inclusion of the structural geological model consistently improves predictive performance across all tested algorithms, with Random Forest [5] achieving the best coefficient of determination (R^2). Spatial generalization is evaluated using a Leave-One-Borehole-Out (LOBO) protocol with increasing spatial exclusion buffers (100m to 10km), designed to disentangle memorization from a fair spatial generalization. SHAP [6] analysis confirms that RGT is among the most influential features, providing interpretable evidence for the added value of geological knowledge integration into the pipeline. We finally discuss the relative contribution of each data modality, the limitations of uncertainty estimates using Gaussian processes [7] and conformal prediction [8] under covariate shift due to spatial correlation [9]. Representative cross-sections illustrate predicted velocities, associated uncertainties, and the influence of geological structure (RGT) on model behavior according to the treeSHAP algorithm [6]. Ongoing developments include the integration of surface geological contacts to improve the structural model, and the application of natural language processing techniques to extract lithological properties (grain size, chemistry) from detailed textual descriptions in explanatory notes of geological maps.



Left: Surface geological age map of the study area. **Right:** Cross-section at fixed longitude with, from top to bottom: RF-predicted S-wave velocities (m/s); 90% confidence interval width, evaluated with a spatial-aware cross-conformal calibration; Relative Geological Time (RGT) voxel model; SHAP values for the RGT feature, its contribution to the overall RF prediction.

- 1 Sørensen K.I., Auken E. A new system for airborne TEM surveying. *Exploration Geophysics*, 35(3), 191–199, 2004.
- 2 Reimers N., Gurevych I. Sentence-BERT: Sentence embeddings using siamese BERT-networks (2019). arXiv:1908.10084.
- 3 Grose L., et al. LoopStructural 1.0: time-aware geological modelling. *Geosci. Model Dev.*, 14, 3915–3937, 2021. doi:10.5194/gmd-14-3915-2021
- 4 Brown, C. F., et al. Alphaearth foundations: An embedding field model for accurate and efficient global mapping from sparse label data. arXiv preprint arXiv:2507.22291.
- 5 Breiman, L. Random Forests. *Machine Learning* 45, 5–32 (2001). doi.org/10.1023/A:1010933404324
- 6 Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.
- 7 Williams, C., & Rasmussen, C. (1995). Gaussian processes for regression. *Advances in neural information processing systems*, 8.
- 8 Cordier, Thibault, et al. “Flexible and systematic uncertainty estimation with conformal prediction via the mapie library.” *Conformal and Probabilistic Prediction with Applications*. PMLR, 2023.
- 9 Tibshirani, R.J., Barber, R.F., Candès, E.J., & Ramdas, A. (2019). Conformal Prediction under Covariate Shift. *Neural Information Processing Systems*.

Automatic Geo-AI Workflow for Stress Information Extraction: Detecting, Digitizing, Harmonizing and Validating Measurements from Legacy Documents for the Stress Database Netherlands

M.D. Demmers¹, Anniek Theuwissen¹, Hamed Ali Diab-Montero¹, Loes Buijze¹, and Jurrien Dijk¹

¹ TNO – Geological Survey of the Netherlands, Utrecht, The Netherlands

Large volumes of subsurface data from the oil and gas domain remain locked in unstructured and semi-structured document formats, including legacy reports, scanned archives, and heterogeneous digital files. Within national subsurface data infrastructures, many of these documents originate from historical operator submissions and paper archives that were never systematically inventoried or digitized, an issue that becomes more pressing as operators exit mature basins and custody of legacy archives changes. The absence of structured inventories and standardized ingestion workflows creates a substantial risk of irreversible data loss and severely limits reuse potential for emerging subsurface applications such as geothermal energy and carbon capture and storage (CCS).

A particularly valuable subset of these materials concerns subsurface stress data, which is essential for safe and optimal drilling and for subsurface safety and integrity. Formation Integrity Tests (FIT) and Leak-Off Tests (LOT) are a key data stream, but their pressure–flow information is typically embedded in tables, plots, and figures across a wide range of technical documents. While this information has traditionally been interpreted manually, the scale and heterogeneity of the document landscape make such approaches increasingly infeasible. Moreover, national-scale, data-driven analysis requires that FIT/LOT measurements are not only extracted, but also harmonized, validated, and stored consistently before any quantitative interpretation can take place.

The objective of this work is to develop and demonstrate an end-to-end automated framework that transforms unstructured subsurface pressure and stress information from heterogeneous documents into a unified, database-ready form, so that results can be appended to the Stress Database Netherlands and used in a consistent way by stakeholders. The focus is on scalable extraction, standardization, and validation of tabular and graphical stress data, thereby creating the necessary foundation for subsequent data-science-driven analysis.

We implement these capabilities in a Human-in-the-Loop (HITL) platform in which data management plays a central role and computer vision (CV) accelerates the process. Relevant objects in unstructured documents are detected using deep-learning-based object detection, and data managers work closely with AI systems: accreditations of detected attributes and corrections are captured as valuable training data to continuously improve the models. The workflow finds relevant objects, digitizes the values within them, standardizes extracted values into defined schemas, validates records against schema constraints and expert bounds, and subsequently supports (automated) interpretation to derive secondary metrics that are commonly used by stakeholders. Object detection is performed with two CV models for locating tables containing pressure–flow measurements and one model for locating figures where pressure–flow curves are plotted. Digitization leverages established approaches: OCR and out-of-the-box tools are used to digitize tables and extract values from graphics, after which standardization fits extracted values per object to the defined schema and automated validation is applied, with anomalies flagged for expert review.

The workflow produces a structured, validated dataset of subsurface pressure measurements extracted from unstructured documents. Each record includes standardized attributes such as well identifier, test type (e.g., FIT or LOT), measured pressures, depths, and provenance metadata (document source, page, and figure/table reference). Tables and figures are digitized into machine-readable formats, extracted values are harmonized according to the schema, and automated validation ensures that values are consistent with expected physical ranges and formats. Overall, the framework provides an end-to-end approach for transforming heterogeneous subsurface documents into a unified, validated dataset of stress measurements, reducing reliance on manual interpretation, mitigating the risk of data loss from legacy archives, and enabling consistent, repeatable workflows that support integration into the Stress Database Netherlands and scalable downstream analyses of wellbore integrity and drilling operations.

A Deep Learning Approach for Automated Vectorization of Historical Mining Maps

David Demmers¹, C.P. van Middelkoop¹, Samantha Kim¹, Joana E. Martins¹, Erik van Linden¹

¹TNO – Geological Survey of the Netherlands, Utrecht, The Netherlands

Historical mining maps hold detailed geospatial information, yet most are currently only archived as static images while holding features that can tell us much more about specific areas. [1] In the Limburg province, in the Netherlands, approximately 3500 historical mining maps document all mining activities, but they have never been vectorized in high detail, nor have depth and temporal information been systematically extracted from the maps.

To extract this data we developed a deep learning workflow for automated extraction of underground mining structures from historical coal mining maps. The core component is a semantic segmentation model based on the U-Net architecture. The model performs pixel-wise classification to delineate two primary feature classes: mining panels and galleries, together with a background class.

The dataset consists of approximately 3,500 georeferenced rasterized maps documenting about 150 years of mining activity in Southern Limburg. The collection, curated by the Geological Survey of the Netherlands - TNO, contains maps produced by multiple private mining operators with varying cartographic notations. All maps had been digitized and georeferenced prior to this study. From this corpus, 2,000 image patches (4096 × 4096 pixels) were manually annotated with semantic masks for panels, galleries, and background. In order to evaluate the trade-off between spatial detail and contextual information the annotated dataset was prepared in three resolutions: the original resolution, half resolution and quarter resolution. Experiments showed a consistently better model fit for both primary feature classes when trained on the half-resolution dataset.

We trained a custom U-Net using the Dice Similarity Coefficient loss on the training data, which consists of 70% of the data set. While monitoring the model performance on unseen validation data using both the Dice coefficient and the Jaccard index. Both metrics are used to ensure a more robust validation. The validation data consists of 21% of our dataset. In addition to supervised validation a separate self-supervised inspection of model outputs is used to identify error rates and uncertainty patterns across the corpus when compared to a rough manual vectorization of previously annotated coal layers.

Evaluation using the Jaccard index (Intersection over Union, IoU) yielded scores of 0.985 for background, 0.896 for panels, and 0.676 for galleries. Corresponding Dice Similarity Coefficient (DSC) scores were 0.993 for background, 0.942 for panels, and 0.803 for galleries. The model demonstrates high segmentation accuracy for the features, while gallery structures do remain comparatively more challenging due to their smaller spatial extent and greater visual variability across maps.

The results demonstrate that the proposed U-Net based workflow can accurately extract mining panels and galleries from heterogeneous historical coal mining maps, producing high-quality raster representations. The workflows generated geometries are suitable for further digitization and spatial analysis, performing well across our diverse corpus spanning multiple operators and conventions.

Overall, the results represent a significant step toward the automated digitization of historical mining maps. Further improvements will likely benefit from training on a broader and more varied dataset to better capture stylistic variation across the corpus. Important attributes such as excavation dates and depth measurements, currently embedded as textual annotations, will be addressed in future work where we leverage Vision Language Model (VLM) based text extraction methods.

Classification of geological borehole descriptions using a domain adapted large language model

Katrijn Dirix¹, Hossein Ghorbanfekr², Pieter Jan Kerstens³

1. VITO [Environmental Intelligence, katrijn.dirix@vito.be]; 2. IMEC; 3. VITO [Data & Analytics]

Geological borehole descriptions contain rich, expert generated textual information that is essential for subsurface modelling, geotechnical applications, mineral exploration, and groundwater research. Yet, their unstructured natural language format poses a major challenge for automated interpretation. Traditional rule-based extraction methods, which rely on dictionaries and handcrafted regular expressions, struggle to capture the linguistic complexity and variability inherent in these descriptions.

In early 2024, we initiated research to explore whether large language models (LLMs) could more effectively extract structured lithological information from free text borehole descriptions. We developed GEOBERTje, a domain adapted Dutch language model derived from BERTje and further trained via masked language modelling on > 280,000 unlabeled geological descriptions. A smaller manually curated dataset of ~2,500 labeled intervals was then used to fine tune a multi label classifier capable of assigning a primary and two secondary lithology classes.

The resulting model demonstrates consistent improvements over three alternatives: a generic BERTje based classifier, traditional rule-based scripts, and a prompt engineering approach using GPT 4. GEOBERTje especially excels in capturing secondary and tertiary lithologies, where linguistic subtleties (e.g. nested descriptions, indirect grain size qualifiers, or multiple co dominant lithologies) often cause existing systems to fail or hallucinate. Unlike GPT 4, the domain adapted model rarely invents lithologies that are not present in the original descriptions, which is crucial for downstream geological modelling workflows.

This work is fully described in Ghorbanfekr et al. (2025), in which the methodology, accuracy benchmarks, and error analysis are presented in detail. For the conference contribution, we will include an updated performance benchmark incorporating GPT 5, allowing us to assess whether domain adaptation remains beneficial for this problem, or general frontier models have now caught up.

Ghorbanfekr, H., Kerstens, P. J., & Dirix, K. (2025). Classification of geological borehole descriptions using a domain adapted large language model. *Applied Computing and Geosciences*, 25, 100229.

Graph-based geological modeling with faults using the geostatistical SPDE approach

Charlie Garayt^{1,2}, Nicolas Desassis¹, Nicolas Clausolles², Simon Lopez²

(1) Mines Paris - PSL

(2) BRGM

Stochastic partial differential equation (SPDE, Lindgren *et al.*, 2011) is a novel approach in geostatistics to generate a Gaussian random field. The solution of the SPDE is obtained on a mesh whose resolution is dependent on the spatial structure of the modeled Gaussian process. Then, solving the SPDE involves a numerical scheme, in our case Finite Elements, with a stencil that connects the locations of interest and can be seen as a graph. Then, the Gaussian distribution at any location in the domain can be approximated using the SPDE solution. Conditioning observations are integrated as constraints expressed as a linear combination of vertex distributions. Our work recasts the potential field method (Lajaunie *et al.*, 1997), an implicit geological modeling method, within the SPDE framework whose originality was to integrate both contact points and structural orientation measurements as modeling constraints.

As faults are major components of geological structural models. The relative displacement of geological layers on both sides of these discontinuity surfaces is challenging to model. Many methods have been proposed in the geological modeling literature. Concerning geostatistical approaches, the correlation between layers across the fault surfaces is neglected (*screen effect*, Marechal, 1984) or relies on discontinuous drifts to model this displacement (Calcagno *et al.*, 2008). We propose to correlate the domains on each side of a fault, extending the natural graph-based formulation that comes with the SPDE approach.

Under the assumption that fault surfaces are known, we create a mesh that conforms to the faults. Then, the mesh is split along fault surfaces, and a virtual mesh is connected to these inner boundaries. In a graph formalism, a sub-graph is added to reconnect duplicated vertices resulting from the splitting operations. Then, applying local anisotropies (Pereira, 2020) to this sub-graph allows correlating points according to the fault displacement.

Calcagno, P., Chilès, J. P., Courrioux, G., & Guillen, A. (2008). Geological modelling from field data and geological knowledge. *Physics of the Earth and Planetary Interiors*, 171(1-4), 147-157. <https://doi.org/10.1016/j.pepi.2008.06.013>

Lajaunie, C., Courrioux, G., & Manuel, L. (1997). Foliation fields and 3D cartography in geology : Principles of a method based on potential interpolation. *Mathematical Geology*, 29(4), 571-584. <https://doi.org/10.1007/BF02775087>

Lindgren, F., Rue, H., & Lindström, J. (2011). An Explicit Link between Gaussian Fields and Gaussian Markov Random Fields : The Stochastic Partial Differential Equation Approach. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 73(4), 423-498. <https://doi.org/10.1111/j.1467-9868.2011.00777.x>

Marechal, A. (1984). Kriging Seismic Data in Presence of Faults. In G. Verly, M. David, A. G. Journel, & A. Marechal (Éds.), *Geostatistics for Natural Resources Characterization* (p. 271-294). Springer Netherlands. https://doi.org/10.1007/978-94-009-3699-7_17

Pereira, M. (2019). Generalized random fields on Riemannian manifolds : Theory and practice (Publication No. 2019PSLEM055) [Theses, Université Paris sciences et lettres]. <https://pastel.hal.science/tel-02499376>

Towards integrated seismic characterisation of fluvial geothermal reservoirs with deep learning

B. S. I. Gesbert¹, D.J. Verschuur¹, G. Song¹, H. A. Abels¹, S. Geiger¹

¹ Geoscience and Engineering Department, Delft University of Technology

Geothermal developments require a high success rate because of their challenging finances. This usually restricts seismic data acquisition to 2D lines (Rehling et al., 2023), which are impractical for accurately interpreting 3D channelised reservoir connectivity (Posamentier et al., 2022), posing considerable uncertainty to geothermal production estimates and well drilling decisions. Given data scarcity and constrained budgets, it is key in geothermal projects to explore a wide range of geological uncertainties and to exploit available seismic data without excessive computational costs.

We build a comprehensive deep-learning-based framework that couples geological scenario modelling with seismic forward modelling and impedance inversion for efficient seismic characterisation of fluvial geothermal systems. The Rapid Reservoir Modelling software is employed, enabling quick probabilistic construction of numerous different 3D sandstone-shale reservoir models (Jacquemyn et al., 2021; Song et al., 2025). A finite-difference modelling workflow leads to a diverse deep learning dataset of 2D seismic depth images, which are inverted for underlying acoustic impedance distributions by training a conventional U-Net architecture (Ronneberger et al., 2015).

Inference of the trained U-Net on various unseen reservoir scenarios shows good inversion outcomes that are remarkably similar between seismic images filtered to high (112 Hz) and low maximum frequencies (40 Hz; Fig. 1). Additionally, applying random seismic noise seems to not matter much above an overall signal-to-noise ratio of around 1 in the image domain, reflecting substantial image distortion. Next steps include further increasing random noise and geological complexity, as well as evaluating how deep-learning inversion predictions affect geothermal flow simulation.

Jacquemyn, C., et al. (2021). Sketch-based interface and modelling of stratigraphy and structure in three dimensions. *Journal of the Geological Society*, 178(4), jgs2020-187.
 Posamentier, H. W., Paumard, V., & Lang, S. C. (2022). Principles of seismic stratigraphy and seismic geomorphology I: Extracting geologic insights from seismic data. *Earth-Science Reviews*, 228, 103963.
 Rehling, J., et al. (2023). The Dutch SCAN Geothermal Exploration Programme—The first three years of 2D seismic acquisition and (re) processing. *First Break*, 41(1), 33-43.
 Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Cham: Springer international publishing.
 Song, G., Geiger, S., Voskov, D., Abels, H. A., & Vardon, P. J. (2025). Scenario-based data assimilation framework to improve production estimates for geologically complex geothermal reservoirs. *Authorea Preprints*.

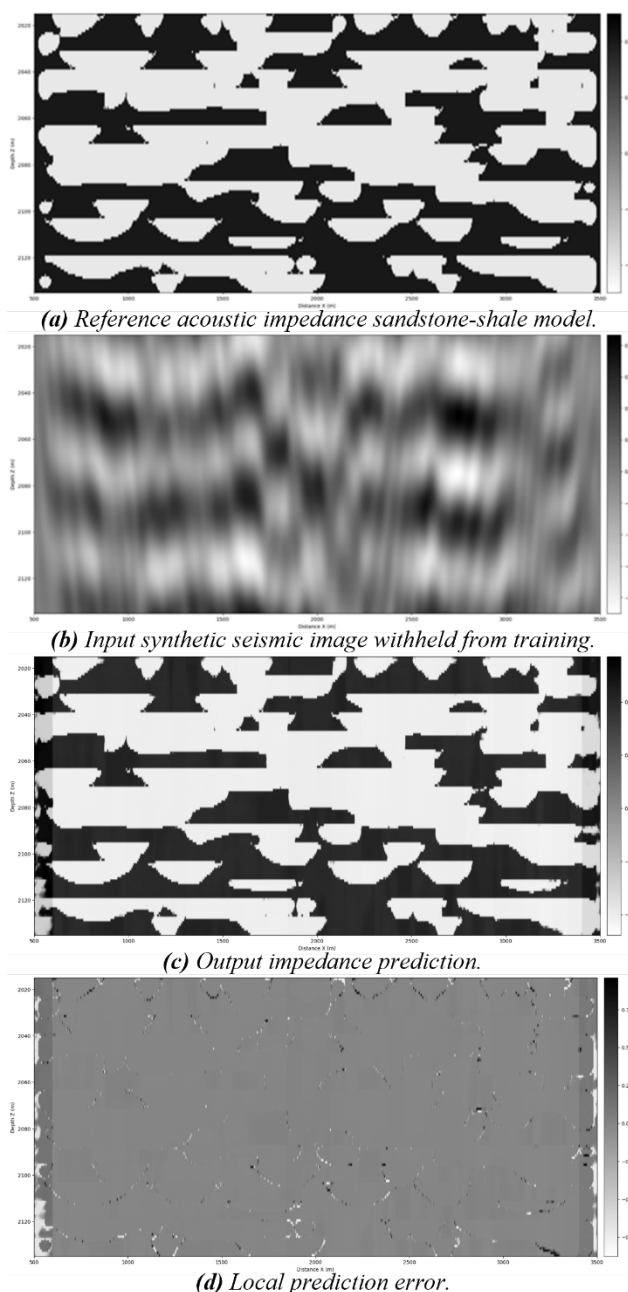


Fig 1. Convolutional neural network (U-Net) inversion results on a synthetic sandstone-shale reservoir test case. The simulated seismic image has been filtered to 40 Hz at maximum, after which it has been perturbed with random noise.

AI for Data Mining and Knowledge Extraction for Mineral Exploration in Greenland

Pierpaolo Guarnieri, Nikolai Andrianov, Jakob Molander

GEUS, Geological Survey of Denmark and Greenland

Mineral exploration in Greenland has generated thousands of geological field reports, mostly stored as scanned PDFs that are usually not searchable. To extract useful geological knowledge from this vast archive at scale, automated methods are needed to process dense technical text, tables, figures, and hand-drawn diagrams within each document.

The geological datasets available for mineral exploration in Greenland are typically multimodal, multidimensional, heterogeneous, complex, and uncertain, with missing values and noise. This preliminary project focuses on a selected area in East Greenland, Jameson Land, which represents an ideal test site due to a large geological and geophysical dataset, a long mineral-exploration history, and well-known mineral occurrences. The aim is to explore how generative AI and ML can leverage existing data to recognize areas with high mineral prospectivity, with a particular focus on key resources, and the identification of hidden patterns in the data.

The AI approach will enable data cleaning and extraction of geological indicators through automated tasks for large-scale screening, reorganizing, and extraction of information from unstructured data, such as scanned PDFs, using a search routine followed by LLM analysis of the text files. The first step consists of document parsing, converting each typed PDF report to a Markdown text file, using an OCR that faithfully reconstructs not only running text but also the spatial semantics of tables, the descriptive content of figures, and the symbolic annotation of diagrams and technical drawings, running on a local GPU. The second step involves analyzing the reports for established keywords associated with a specific mineral system and counting hits, generating georeferenced locality names, and finding localities that co-occur with attribute keywords in the same paragraph and producing a list of candidates for further review. The third step follows the LLM confirmation, partitioning parsed documents into chunks that are semantically coherent, contextually self-contained, and optimally sized for downstream embedding. During this step, an LLM reads the relevant text passages and decides whether each recognized locality is genuinely associated with the established keywords.

Preliminary results from reports covering Jameson Land suggest that co-occurrence of locality names with mineral-system keywords provides a tractable first pass screening signal, substantially reducing the manual review burden. The conference contribution will present the implemented workflow, an assessment of retrieval quality on a test document set, and a discussion of failure modes specific to heterogeneous geological report collections. If validated, the approach is directly transferable to other frontier regions with comparable legacy data archives.

Raster Structure Encoding: Computer Vision Methods for Geophysical Feature Engineering

Tyler Howe¹, Christian Ortmann¹

¹MinersAI

Many traditional machine learning (ML) methods used in the geosciences ingest gridded data as collections of spatially independent pixels, effectively prohibiting models from capturing the structural and textural context critical to geological interpretation. To address this, a modular feature engineering framework is presented for reduced-to-pole (RTP) magnetic data that encodes spatial, structural, and directional information into a multi-channel feature tensor compatible with standard pixel-wise ML workflows. The pipeline first applies local contrast normalization (LCN) [1] and global normalization to the input raster, followed by a signed Gaussian-difference bandpass decomposition [2] that separates the data into complementary high- and low-frequency responses corresponding to fault-like edges and large-scale bodies, respectively, while preserving polarity for positive/negative half-wave separation. From these responses, ten structured feature products are computed across three families: fault and structure features, including Fault Persistence [3] and Fault Density; body and regional features, including a Body-to-Fault Ratio, Multi-Scale Coherence and Coherence Density [4] derived from structure tensor [5, 6] analysis; and fabric and characteristic features, including a sin/cos-encoded Fabric Orientation and Characteristic Region Labels from unsupervised clustering of the full feature set.

Applied to RTP magnetics data from the Arabian Shield, the resulting feature products show strong consistency with manual geological interpretation, clearly delineating NW-trending fault corridors, dense fracture zones, plutonic bodies, and regions of directional fabric alignment. The framework operates entirely on the normalized input raster at native grid resolution, and as such is model-agnostic, requiring no labelled training data for feature generation. Its modular design supports adaptation to other raster types and geological settings, offering a repeatable, systematic preprocessing workflow with potential to improve the sensitivity of prospectivity models to structural controls. Future work will include quantitative evaluation across multiple datasets and geological domains, and assessment of downstream impacts on ML model performance.

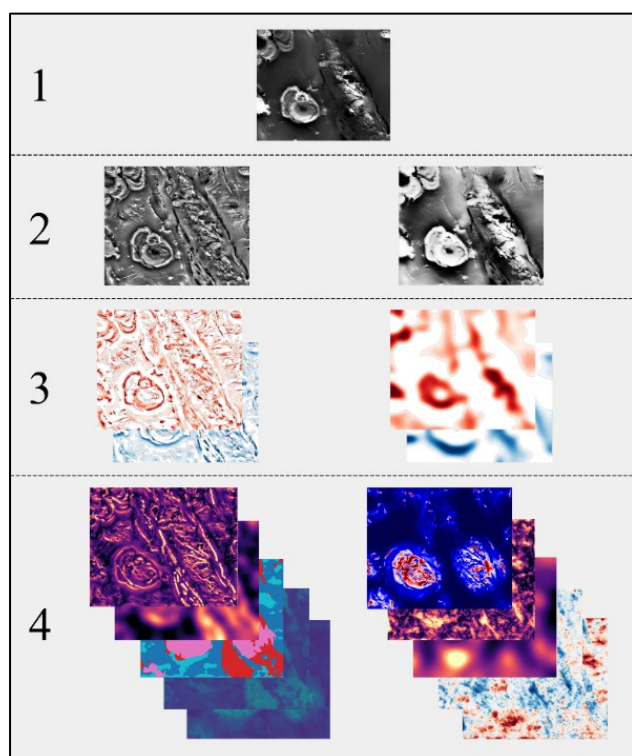


Figure 1. Overview schematic of workflow. Top to bottom: (1) input raster, (2) LCN and globally normalized rasters, (3) multi-frequency signed bandpass outputs, (4) resulting feature set.

[1] N. Pinto, D. D. Cox, and J. J. DiCarlo, "Why is real-world visual object recognition hard?" *PLOS Computational Biology*, vol. 4, no. 1, p. e27, Jan. 2008, doi: 10.1371/journal.pcbi.0040027.

[2] D. Marr and E. Hildreth, "Theory of edge detection," *Proceedings of the Royal Society of London. Series B, Biological Sciences*, vol. 207, no. 1167, pp. 187–217, 1980.

[3] Y. Sato, S. Nakajima, N. Shiraga, H. Atsumi, S. Yoshida, T. Koller, G. Gerig, and R. Kikinis, "Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images," *Medical Image Analysis*, vol. 2, no. 2, pp. 143–168, Jun. 1998, doi: 10.1016/S1361-8415(98)80009-1.

[4] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *Medical Image Computing and Computer-Assisted Intervention — MICCAI'98*, W. M. Wells, A. Colchester, and S. Delp, Eds., Lecture Notes in Computer Science, vol. 1496. Berlin, Germany: Springer, 1998, pp. 130–137, doi: 10.1007/BFb0056195.

[5] J. Bigun and G. H. Granlund, "Optimal orientation detection of linear symmetry," in *Proceedings of the IEEE First International Conference on Computer Vision*, London, U.K., Jun. 1987, pp. 433–438.

[6] T. Lindeberg, "Feature detection with automatic scale selection," *International Journal of Computer Vision*, vol. 30, no. 2, pp. 79–116, 1998. [Technical report ISRN KTH/NA/P-96/18-SE, KTH Royal Institute of Technology, 1996.]

Machine-learning segmentation of ERT and DEM-of-Difference analysis for quantitative landslide activity modelling: a case study from the Dynów Foothills (southern Poland)

Mirośław Kamiński,

Polish Geological Institute-NRI, 00-975 Warszawa, Rakowiecka 4

This study presents an integrated, multi-sensor workflow for quantitative landslide activity modelling that combines photogrammetric topography reconstruction, airborne laser scanning (ALS) and electrical resistivity tomography (ERT) with machine-learning-assisted interpretation. The analysis is based on three input terrain models: a digital terrain model derived from archival black-and-white aerial photographs acquired in 2002 (scale 1:10,000) and two ALS LiDAR-based terrain models from 2013 and 2022. DEMs of Difference (DoD) were computed for 2002–2013 and 2002–2022 to quantify spatial patterns of elevation change, vertical displacements and volumetric mass redistribution. The DoD results indicate a total mobilized colluvial volume exceeding 30,000 m³ and reveal distinct depletion and accumulation zones consistent with landslide kinematics. Subsurface conditions were characterized using an ERT profile extending from the main scarp across the entire landslide body. The resistivity model delineates the geometry of the shear zone to depths of ~10 m and highlights lithological and hydrogeological contrasts, calibrated against four archival boreholes. To reduce interpreter subjectivity and enable reproducible zoning of resistivity domains, ERT data were classified using unsupervised k-means clustering and supervised Random Forest modelling trained on borehole-derived reference information. The supervised approach further provides probabilistic class assignments that help identify ambiguous zones and support robust integration with geomorphic evidence. Finally, the reconstructed activity phases were compared with rainfall records, linking observed deformation patterns with hydro-meteorological forcing. The proposed methodology offers a transferable template for landslide mapping and long-term monitoring in flysch terrains by coupling surface-change quantification with ML-supported subsurface segmentation.

Beyond classified geological maps: the case for AI-enabled mapping of geological properties

Charlie Kirkwood

University of Exeter Centre for Environmental Intelligence

Traditional geological maps represent our knowledge of Earth's geology and are constructed by classifying the ground into discrete units (polygons in 2D, volumes in 3D) based on changes in properties such as age, composition, and texture. These traditional classified maps offer a summary of regional geology, but – in addition to being unable to represent gradational changes – the classified nature of these maps poses a problem for their verification, because the most appropriate class label for any location cannot be objectively measured, only subjectively asserted.

It is therefore inevitable that different geologists will draw different class boundaries in different locations according to their differing judgement, even if hypothetically they may agree entirely on the physical and chemical properties of the ground itself – which are objectively measurable. This conflation of objectively measurable properties with subjectively interpreted classes makes it impossible to definitively assess whether one classified map is closer to reality than another, and therefore also whether proposed updates genuinely improve accuracy, or just better match the proposer's interpretation. This situation makes progress difficult and continues to cause issues with inconsistency across map-tile and political boundaries where different geologist's interpretations meet.

We propose a simple but fundamental shift in approach to separate the objectively measurable from the subjectively inferred, and therefore avoid their conflation. We propose that geological mapping should be conducted as a two-stage process:

1) Model and map predictively, as continuous spatial fields, the objectively measurable properties on which our traditional classifications are based, as well as any other properties of interest (e.g. mechanical properties, critical metal concentrations, geothermal heat flow, presence of fossils etc). These maps of objectively measurable properties, which AI-techniques can be designed to produce, can be objectively verified by comparing their predictions to real measurements of these properties. Better maps will close the gap between predictions and observations in held-out test data. 2) For any given predictive map of objectively measurable properties, proceed if necessary to segment and classify this map output into different units, e.g. of lithology, lithostratigraphy, biostratigraphy, chronostratigraphy etc. We note that A) this discretisation step will inherently result in loss of information to quantisation error, unless an infinite number of classes are used (so it is important to retain the underlying property maps). B) this loss of information may be beneficial to aid communication – it is easier to talk about maps composed of named units than continuous but anonymous maps of geological properties. C) the class or unit we choose to assign a location to does not change the objectively measurable geological properties at that location.



Figure 1- Left of diagonal: Bayesian deep-learned predictive map of geological properties, in this case chemical composition with a ternary colour scheme (K - red, Fe - green, Ca - blue). Posterior predictive distributions can be obtained for any location. From <https://geoscientist.online/sections/features/geological-mapping-in-the-age-of-artificial-intelligence/>
Right of diagonal: Traditional classified geological map, in this case British Geological Survey Bedrock Geology. From <https://geologyviewer.bgs.ac.uk>

SemTrack: A Semantic Knowledge Graph Approach to Geological Information Integration

Imadeddine Laouici¹

¹ French Geological Survey (BRGM)

The understanding of the subsurface relies on integrating heterogeneous geological information originating from geological maps, geological models, and textual sources such as reports and scientific publications. In current practice, these sources remain rarely homogenized and are reconciled manually by domain experts, mostly in the context of 3D geomodel construction projects. Even when information is reconciled, existing methods offer limited support for expert knowledge integration, traceability of interpretations, and automated wholistic consistency checking.

We propose SemTrack, an ontology-based integration approach designed to formalize, reconcile, and exploit multi-source geological information within a unified knowledge graph. In this framework, SemTrack integrates structured information extracted from maps and numerical geological models with unstructured knowledge derived from textual documents, all aligned through a dedicated modeling ontology. The resulting knowledge graph supports logical reasoning and knowledge inference using SWRL rules to ensure the consistency of geological constraints and allows to explicitly encode expert interpretations record. This enables the automation of conceptual inconsistencies detection, transparent inference of implicit geological relationships, the completion of missing information across multiple sources, and advanced complex querying of initially heterogenous geological information.

Discrete Morse Theory for predictive mapping from regional geophysical data

Benjamin Le Teurnier, Jihad Bourassi, Pierre-Alexandre Reninger, Aurelie Peyrefitte, Frederic Lacquement

BRGM, French Geological Survey

Interpretation of regional geophysical data most often involves the precise delineation of areas exhibiting contrasting signals to understand the subsurface structure. At the regional scale, we lack labeled data for the use of supervised machine learning methods. Thus, traditional unsupervised clustering methods are usually employed to address this problem [1], [2], [3]. Those methods are not designed to detect reliable transitions within the data. With this in mind, we developed a methodology based on Topological Data Analysis (TDA) [4] to automatically detect delineations, open or closed, that separate the topological features present in the data [5], [6]. Our approach is based on discrete Morse theory to identify critical points and model their relationships within a topological complex, providing a robust representation of the data structure. At the same time, we also apply this analysis to criteria such as the gradient norm, allowing us to trace paths following the greatest spatial variation in the geophysical data. This enables us to associate a shape to the topological features. In order to select the most significant delineations, we hierarchically merge the delimited areas with their neighbours using an ordering that takes into account the topological persistence and the local contrast.

This approach offers an alternative to classical segmentation that is easier to interpret and more robust thanks to the use of topological analysis to guide the extraction of structures and shape in regional geophysical data.

[1] Y. Wang, A. K. Ksienzyk, M. Liu, et M. Brönnner, « Multigeophysical data integration using cluster analysis: assisting geological mapping in Trøndelag, Mid-Norway », *Geophysical Journal International*, vol. 225, no 2, p. 1142-1157, mars 2021, doi: 10.1093/gji/ggaa571.

[2] G. Martelet, C. Truffert, B. Tourlière, P. Ledru, et J. Perrin, « Classifying airborne radiometry data with Agglomerative Hierarchical Clustering: A tool for geological mapping in context of rainforest (French Guiana) », *International Journal of Applied Earth Observation and Geoinformation*, vol. 8, no 3, p. 208-223, sept. 2006, doi: 10.1016/j.jag.2005.09.003.

[3] D. Eberle, D. Hutchins, S. Das, A. Majumdar, et H. Paasche, « Automated pattern recognition to support geological mapping and exploration target generation – A case study from southern Namibia », *Journal of African Earth Sciences*, vol. 106, p. 60-74, juin 2015, doi: 10.1016/j.jafrearsci.2015.03.011.

[4] G. Rote et G. Vegter, « Computational Topology: An Introduction », in *Effective Computational Geometry for Curves and Surfaces*, J.-D. Boissonnat et M. Teillaud, Éd., Springer Berlin Heidelberg, 2006, p. 277-312. doi: 10.1007/978-3-540-33259-6_7.

[5] T. Ophelders, A. Schenfisch, W. Sonke, et B. Speckmann, « Computing Geomorphologically Salient Networks via Discrete Morse Theory », *LIPICs*, Volume 332, *SoCG 2025*, vol. 332, p. 70:1-70:17, 2025, doi: 10.4230/LIPICS.SOCG.2025.70.

[6] J. P. Bullenkamp, T. Kaiser, F. Linsel, S. Krömker, et H. Mara, « Discrete Morse theory segmentation on high-resolution 3D lithic artifacts », *it - Information Technology*, vol. 66, no 6, p. 191-205, oct. 2024, doi: 10.1515/itit-2023-0027.

AI Knowledge Navigator for Norwegian Offshore Data

Jesse Lord ¹, Petter Dischington ²

¹ FabriqAI AS, ² Norwegian Offshore Directorate

The Norwegian Offshore Directorate, together with FabriqAI, is creating the Knowledge Navigator, a context-aware discovery portal for evidence-grounded synthesis across heterogeneous offshore data sources. The ambition goes beyond document retrieval; users of various levels of expertise converse in natural language with an assistant to either answer the question or retrieve the data required.

Subsurface, environmental and operational knowledge are available for interpretation and reuse across maps, reports, metadata, tables and figures. The tool is designed for different user groups with different information needs, from a curious citizen wanting to understand where exploration is planned, to an asset manager needing efficient access to regulatory requirements and supporting documentation.

The approach combines semantic search, geospatial filtering, toponym recognition and source-aware ranking to interpret user intent in its geological and geographic context. Queries such as reservoir quality near a specific well, emissions within a defined offshore area, drilling problems in a given formation, APA rounds in a specific region, or regulatory requirements linked to a project phase are translated into structured search plans that coordinate retrieval across multiple portals and document types. A key feature is the ability to ask clarifying questions when user queries are ambiguous or incomplete, allowing the system to refine scope and improve relevance before generating results. Relevant passages, figures and tabulated observations are linked back to their original sources.

Traceability is a core design principle. Rather than returning opaque summaries, the Knowledge Navigator shows where information comes from at the level of document, page, section and geographic object. This supports user validation and makes the output more useful in settings where transparency and accountability matter. The map-based interface also plays a practical role, since many offshore questions depend on location and on how names, areas and geological concepts are defined across sources. We present the tool in its current stage of development and discuss how geo-AI can connect unstructured documentation with spatially referenced knowledge in offshore workflows. The work points toward a new way of combining generative AI with distributed data sources, document repositories and map-based interfaces. It demonstrates how trustworthy AI, federated search and geospatial reasoning can support more efficient interpretation, clearer access to regulatory and technical knowledge, and better reuse of legacy information.

Multi-source integration of geoscientific and economic-mineral knowledge for ranking in mineral exploration

P. X. P. Martin-Cocher

Infomineral

This work proposes the systematic integration of multi-source geoscientific and economic-mineral knowledge assisted by AI, an attempt to make geology more continuous and less discrete in its boundaries.

The framework ingests multiple heterogeneous sources as weighted layers of a single inference system, without manual pre-processing, forced nomenclature standardisation, or sequential steps between domains. Geological sources include: maps and explanatory notes from geological surveys; scientific articles and theses; more than 15,000 technical mineral exploration reports (NI 43-101, JORC, SK-1300) processed by LLM-based pipelines used as data extraction tools; and petrogenetic compilations. Using epistemological weighting mechanism, the system incorporates economic-mineral context: mining legislation and blocked areas; environmental protection zones; political situation and resource sovereignty; commodity price history and projections; profile of active investors in the region; and economic viability parameters (grade cut-off, minimum deposit size, investigation CAPEX). Each source receives weight proportional to its scale, purpose, and evidential quality. The result is a continuous probability field over lithological subtypes and mineral context (suggested name hypergeology), replacing discrete polygons. In practical terms, the unit of evaluation becomes the polygon or mineral title, organised through a ranking system.

The framework presents three key differentials. First, adaptive lithological hierarchy: in data-dense zones (e.g., Borborema Province, NE Brazil), the model descends the arborescent structure, "granite" expands into two-mica granite, rapakivi granite, and associations such as pegmatite and aplite. In data-sparse zones (e.g., Nigerian mobile belts, with analogous geology but far more limited documentation), the model ascends one or two levels (operating on "granitoid") without imposing unsupported subdivisions. Granularity is determined by local observation density, allowing coherent operation across regions with asymmetric data coverage. Second, practical falsifiability: when an area is flagged as favourable, Pit Detect is executed, a routine developed by the author, trained on satellite images of pits derived from OpenStreetMap polygons, capable of identifying artisanal, semi-artisanal, and industrial mining activity. Artisanal miners are historically efficient and low-cost prospectors; their presence is empirical evidence of attractiveness, and their absence in a predicted favourable area is a significant (though not definitive) signal of model failure. Third, investment-decision output: the result is not "where something might exist", but where it is worth investigating, for which investor profile, with which minimum budget, and (critically) what to do to try to break the model, not to confirm it.

The final output includes: a comparative ranking of targets within a budget window; the appropriate investor profile (early-stage explorer, major, State); regulatory and political conditions required for acceptable risk; and the minimum field dataset needed to validate or refute the model's predictions. For geological surveys, the framework enriches legacy vector archives with uncertainty layers and economic context without remapping; for explorers and investors, it concentrates effort where geological probability, economic viability, and the regulatory environment converge.

The model is in a development, evaluation, and testing phase. Results for specific commodities and areas are underway, but the framework does not yet constitute a global or multi-commodity model. This work describes the conceptual architecture, methodological principles, and available preliminary results, with the aim of opening the framework to community discussion and contribution.

Process-informed spatial prediction for soil mapping: towards hybrid modelling frameworks for Earth science applications

Marijn van der Meij¹, Madlene Nussbaum¹

¹ Department of Physical Geography, Utrecht University, The Netherlands

Recent advances in artificial intelligence (AI) and machine learning have led to rapid progress in spatial prediction and mapping across the geosciences. Data-driven approaches, often based on Earth observation data and large environmental covariate sets, have demonstrated strong predictive performance in a wide range of applications, including geological mapping, mineral exploration, and soil property prediction.

However, these approaches also face fundamental limitations, particularly when applied in situations characterized by sparse observations, complex processes, and strong spatial heterogeneity. Especially for soil maps, predictions created by data-driven machine learning often suffer from inconsistencies. Interrelated soil properties showing typical vertical distributions are predicted one-by-one resulting in inconsistent sets of properties. Such inconsistencies introduce significant uncertainties into subsequent map applications and can lead to unrealistic outcomes, particularly when dynamic processes are wrongly specified due to unrealistic soil inputs.

This project proposes the development of a process-informed machine learning framework that integrates mechanistic knowledge with advanced machine learning techniques to generate consistent, high-resolution global soil maps. To provide global process-information to the machine learning, an application oriented numerical soil classification is developed based on new similarity metrics to compare full soil profiles. This results in a soil profile reference library, which is verified and expanded with expert knowledge. The reference library then informs two novel ML approaches for global space-time mapping: loss function-constrained and transformer-based machine learning. The newly created global soil maps are used as inputs to land surface model to simulate biomass influenced by soil nutrient dynamics. The effect of existing inconsistent and the new consistent soil maps are benchmarked.

By combining process-informed modeling with advanced uncertainty quantification, the project aims to move beyond purely data-driven GeoAI toward approaches that are both scientifically grounded and operationally useful, particularly relevant in contexts where data are limited and process understanding remains essential.

A pragmatic theoretical framework for geomodelling in hyperspace

Hilco van Moerkerk¹ and Milan Milanov²

1. GEOREKA Technologies GmbH, Austria; 2. Octree Consulting, Australia

Most Implicit Modelling (IM) techniques are fundamentally the same as dual Kriging. Horowitz et al. already wrote about this mathematical similarity for Radial Basis Functions (RBFs)^[1]. We already extended that mathematical foundation to explain further similarity to Support Vector Machines (SVMs) and Single Layer RBF networks in a blog post (not peer reviewed).

The problem is that these methods require fundamental statistical parameters not directly related to the geological structure, such as range. For complex folded geometries those values can be largely meaningless as they crosscut geometry. To improve reliability, one needs to apply complex unfolding, or apply pseudo-methods such as structural trends that follow the geometry but do not account for grade trends. We propose to re-think these interpolations with lessons learned from Machine Learning to improve modelling of complex geology through non-linear mapping to hyperspace. In our work we focused on SVMs, but the theory can be applied to other methods.

In IM signed distance or indicator interpolation is used with values on either side of a contact receiving opposite signs and contact points having a zero distance. These samples are used to determine a set of weights from a Covariance or Kernel Matrix. Estimates are then calculated via the generic formula: $F(x) = \sum(\omega * \varphi(x)) + b$, where the $\varphi(x)$ is a covariance function between the unknown location and the support vectors. In all cases, ω are interpolation weights derived from the matrix and b is a bias term. We can see that this is similar to a high-dimensional hyperspace defined by the $\varphi(x)$ vectors and where the weights vector acts as a normal vector to a 'decision' hyperplane. The contact points of a continuous geometry in 3D space are all mapped to lie on that hyperplane.

The implication is that after modelling the geological structure, other attributes such as grades, can be mapped to, and interpolated within, that geometric hyperspace. Assuming those other parameters are correlated to the underlying geological structure, this means anisotropy can be relatively easily decomposed as tangential to the hyperplane. Analysis of the second stage covariance can be done via experimental variograms where distances are measured in hyperspace instead of 3D. This allows for interpolation within this hyperspace along the structure of the original geometry.

The advantage of this approach is that it can be applied to any initial interpolation. Intuitively, one would first interpolate the main controlling structure, before applying secondary modelling within the resulting hyperspace. However, the method is not constrained to geometry and one could first create a hyperspace on grade, or alteration before applying secondary analysis. Moreover, once a hyperspace is created it can be applied to various other attributes, or the process can even be repeated: map to geometry for structure, re-map to alteration, then apply analysis and interpolation on grade. This allows for an intuitive and flexible chain of mappings from youngest to oldest events adapted to the project without requiring re-training.

This approach is still undergoing real world testing. In a first trial, we applied the mapping followed by simple cluster analysis. The mineral grade of interest was first interpolated using isotropic interpolation. The resulting hyperspace was then used to map geochemistry information and define clusters of high, medium and low grades. Subsequently, all points were compared to the 3 centers in hyperspace and their hyperspace distances rescaled to relative values (0-100%). We could successfully correlate high grades to particular mineralization, which enables more targeted and reliable resource estimation.

Quantifying Subjectivity in Geological Mapping: A Rock Identification Confusion Matrix for Uncertainty-Aware Geo-AI

Anna K. Ksienzyk¹, Jalil Asadi^{2,3}, Bernard Bingen¹, Terje Bjerkgård¹, Morgan Ganerød¹, Deta Gasser^{1,4}, Tor Grenne¹, Ida H. Gunleiksrud¹, Johannes Jakob¹, Gurli B. Meyer¹, Hanne-Kristin Paulsen¹, Anne K. Svendby¹, Alena Bleicher⁵, Hendrik Paasche²

¹ Geological Survey of Norway (NGU), Leiv Eirikssons vei 39, 7040 Trondheim, Norway; ² UFZ – Helmholtz Centre for Environmental Research GmbH, Department Monitoring and Exploration Technologies, Permoserstr. 15, 04318 Leipzig, Germany; ³ now at: Zoi TechCon GmbH; ⁴ Department of Civil Engineering and Environmental Sciences, Western Norway University of Applied Sciences, Røyrgata 6, 6856 Sogndal, Norway; ⁵ Harz University of Applied Sciences, Wernigerode, Germany.

Geological maps are fundamental information layers used in mineral exploration, hazard assessment, land-use planning, and geoscientific research. However, the assignment of rock labels during geological field mapping is inherently interpretative. While field descriptions aim to document observations as objectively as possible, the final classification of a rock type represents a conceptual interpretation based on visual perception and expert judgement [1]. These interpretations introduce subjectivity that is rarely quantified. This subjectivity can be considered as uncertainty that becomes increasingly relevant when geological field observations or derived geological maps are integrated with large digital data sets using machine learning approaches. If the uncertainty associated to geological rock label assignments is not realistically quantified, reliable information retrieval from geological maps becomes impossible. Humans compensate for missing information, such as unquantified uncertainty, associated with data, using experience and intuition. AI systems cannot. For them, quantified data uncertainty is not optional: it is essential for correct information retrieval from data.

We present a structured methodology to quantify human-induced uncertainty in field-based rock identification through a crowd-based expert assessment [2]. Eleven geologists participating in a bedrock mapping project in the Trøndelag region of Norway evaluated the likelihood of confusing pairs of rock types using a five-level Likert scale. From these responses we construct a probabilistic rock identification confusion (RIC) matrix describing the collective uncertainty of the expert cohort across 35 lithological classes. The matrix captures the relational structure of classification ambiguity between rock types and provides a machine-readable representation of expert-derived uncertainty.

We further outline how the resulting probability distributions can be translated into stochastic misclassification cost matrices for loss-sensitive machine learning. Within a Monte Carlo framework, these cost matrices can be sampled during model training, enabling the propagation of expert-derived rock label uncertainty through geological data integration workflows and resulting in probabilistic geological maps.

The presented approach provides a practical pathway to incorporate tacit expert knowledge into uncertainty-aware geo-AI pipelines. While demonstrated for a regional bedrock mapping project, the method is scalable to larger geological surveys and transferable to other geoscientific field classification tasks that rely on human visual interpretation. It therefore contributes to bridging the gap between expert-driven geological mapping practice and emerging AI-based modelling workflows.

[1] Wilson, C.G., Bond, C.E., Shipley, T.F. (2019) How can geologic decision making under uncertainty be improved?. *Solid Earth* 10, 1469–1488.

[2] Ksienzyk, A.K., Asadi, J., Bingen, B., Bjerkgård, T., Ganerød, M., Gasser, D., Grenne, T., Gunleiksrud, I.H., Jakob, J., Meyer, G.B., Paulsen, H.-K., Svendby, A.K., Bleicher, A., Paasche, H. (submitted) Uncertainty in geological mapping: a rock identification confusion matrix for a bedrock mapping project in Trøndelag, Norway. *Norwegian Journal of Geology*.

Towards model-based quaternary geomorphological mapping in Norway: one thematic map at the time

Sven-Arne Quist¹, Mikis van Boeckel¹, Emil Jordhøy¹, Markus Diesing¹, Lena Rubensdotter^{1,2}, Max Holthuis^{1,3}, Jose Santiago Pullarello¹

1. Department of Earth Surface & Seabed, Norwegian Geological Survey (NGU), 7040 Trondheim, Norway

2. Department of Arctic Geology, University Centre of Svalbard (UNIS), 9171 Longyearbyen, Norway

3. Department of geography and social anthropology, Norwegian University of Science and Technology, 7491 Trondheim, Norway

The mounting costs of geohazard-induced damage and the rising demand for mineral resources increase the need for reliable geological maps in Norway. However, due to the time-consuming mapping process, the wide diversity in landscapes and the large areas to cover, nationwide coverage of detailed maps remains out of sight. Especially geomorphological mapping relies heavily on manual interpretation of digital terrain models (DTMs), a step that can be subdivided and enhanced by machine learning. Ideally, this approach results in various model-based nationwide thematic maps, with examples being general sediment distribution, sediment thickness, and landforms.

In this study, sediment distribution and alluvially dominated fan-landforms are modeled using a UNet-based semantic segmentation approach with training data from both legacy databases and newly labeled samples. Sediment thickness was modeled using a pixel-based regression random forest trained on legacy point observations from various databases, coupled with the sediment distribution map and several DTM derivatives as covariates. While the models are rooted in conventional methods, this contribution aims for an honest, data-centric discussion on recurring issues in the development of such thematic maps. This includes, among others, the effect of incomplete training data, label ambiguities, spatial resolution, and the use of legacy versus newly sampled data on model training and their outcomes. By addressing these methodological issues with their potential solution, we share a practical perspective on ongoing model development for geological mapping in general and geomorphological mapping in particular.

Expert-Guided Axiomatic AI Protocol for Facilitating Subsurface Conceptual Modeling

Jose Dario Rodriguez¹, Kris Piessens¹, Kris Welkenhuysen¹, and Thomas Hermans²

1. Geological Survey of Belgium, Royal Belgian Institute of Natural Sciences (RBINS), Brussels, Belgium; 2. Ghent University (UGent), Ghent, Belgium

The initialization of potential-field implicit geological models and dynamic hydrogeological simulations relies heavily on the rigorous parameterization of spatial constraints (Wellmann & Caumon, 2018).

However, translating qualitative, unstructured legacy data, such as drill logs and historical stratigraphic reports, into quantitative, topologically valid 3D inputs remains a significant industry bottleneck.

Conventional Natural Language Processing (NLP) techniques frequently fail to capture the complex spatial reasoning required for layered subsurface modeling. Recent advances have demonstrated that LLMs, when specialized for domain-specific scientific tasks, can substantially outperform general-purpose NLP approaches in knowledge extraction and reasoning within technical fields (Lin et al., 2024). To bridge this semantic-geometric gap, this study introduces an integrated computational pipeline that couples LLMs with a rigid formal ontology, the Lithotectonic Framework (LTF), a structured rule-based schema developed as part of this work, to support and streamline conceptual model generation.

Rather than treating the LLM as a mere text parser, our protocol deploys it as a structured inference engine constrained by formal ontological rules. The workflow operates in three distinct phases. First, during semantic mapping, the LLM dynamically extracts depth intervals, lithological signatures, and hydrogeological properties from raw textual datasets. Second, the system applies LTF axioms, which prioritize boundaries as the defining features of geological units. By constraining the LLM's output to these strict rules, including constraint-based output filtering and schema validation, the system enforces mathematical and topological consistency, substantially reducing probabilistic inference errors and resolving complex stratigraphic ambiguities. Third, before reaching the human expert, the LLM generates a draft layered conceptual framework that undergoes an internal automated audit via structured self-critique.

The draft is then presented to the geologist through a Human-in-the-Loop (HITL) transparency dashboard. This interface tracks metric uncertainties, data provenance, and underlying assumptions, allowing the expert to actively test and refine the interpretation. This approach substantially reduces the data processing burden associated with legacy datasets while ensuring that final geological interpretations remain under expert human control via defined, auditable protocols. The resulting structured outputs are directly translatable into the coordinate geometries required by 3D interpolation engines. Ultimately, this workflow provides robust foundational data that significantly accelerates the development of layered conceptual models, streamlining their preparation as inputs for complex groundwater flow, geothermal energy, and subsurface storage (CO₂, hydrogen, and natural gas) simulations (Kaufmann & Martin, 2008).

Lin, Z., et al. (2024). "GeoGalactica: A Scientific Large Language Model in Geoscience." arXiv preprint.

<https://doi.org/10.48550/arXiv.2401.00434>

Wellmann, F., & Caumon, G. (2018). "3-D Structural geological models: Concepts, methods, and uncertainties." *Advances in Geophysics*, 59, 1–121. <https://doi.org/10.1016/bs.agph.2018.09.001>

Kaufmann, O., & Martin, T. (2008). "3D geological modelling from boreholes, cross-sections and geological maps, application over former natural gas storages in coal mines." *Computers & Geosciences*, 34(3), 278–290. <https://doi.org/10.1016/j.cageo.2007.09.005>

Deep generative modeling of non-stationary sedimentary deposits in 3D

Guillaume Rongier¹, Luk Peeters²

¹ Department of Geoscience & Engineering, Delft University of Technology, Delft, The Netherlands; ² VITO Digital Water & Soils, Mol, Belgium

Subsurface decision-making is hampered by high uncertainties due to sparse data. Embedding geological knowledge into predictive models is therefore essential for robust decisions. Traditional geostatistical methods still struggle to reproduce the architecture and heterogeneity of sedimentary deposits, while recent advances in deep generative models offer a promising alternative, although important challenges remain. The first challenge is obtaining the large number of 3D training images required by such models. To address this, we used a process-based model – the Channel-Hillslope Integrated Landscape Development (CHILD) model – to simulate fluvial evolution over geological timescales (Rongier & Peeters, 2025a). We generated 20,000 simulations of meandering deposits and used them to train a generative adversarial network (GAN). An ablation study shows that techniques developed for generating large 2D images transfer well to 3D fluvial deposits: training remains stable (figure 1) and the GAN generates diverse, non-stationary deposits with detailed internal structures, without mode collapse or memorization. The second challenge is generating realizations conditioned on well and geophysical data. We evaluated several GAN inversion approaches for fitting well and seismic data and found that the GAN latent space is often entangled, making conditioning difficult (Rongier & Peeters, 2025b). But updating the model's weights during the inversion significantly improves the fit to the data. These results suggest that that we are close to being able to use deep generative models in practical applications. But further work is needed to build geologically plausible and diverse training datasets that support the development of foundational generative models, which can then be fine-tuned to specific case studies and conditioning data.

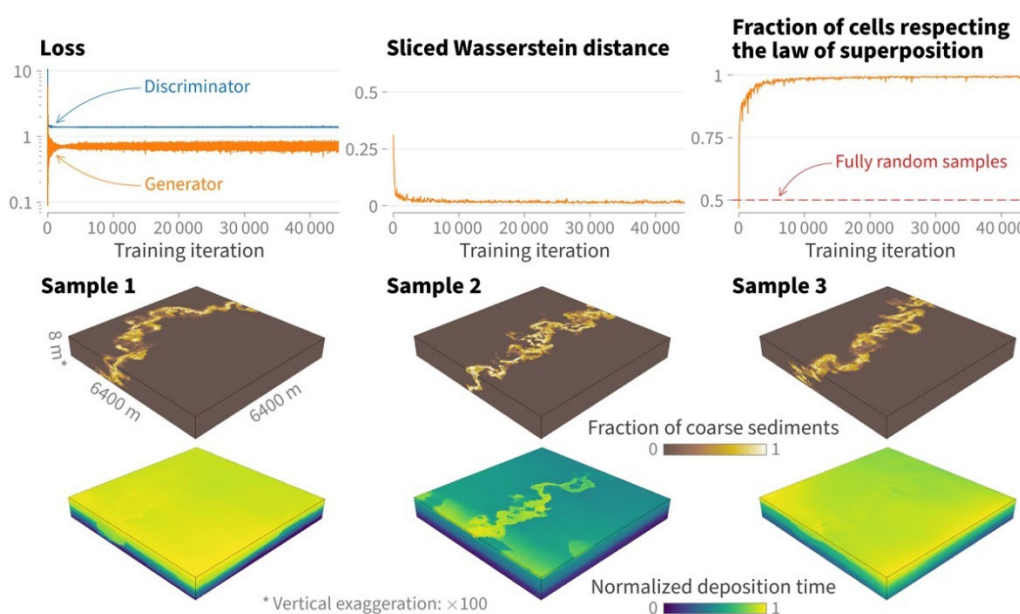


Figure 2. Training a GAN to generate 3D non-stationary fluvial deposits.

Rongier, G., & Peeters, L. (2025a). Towards geological inference with process-based and deep generative modeling, part 1: Training on fluvial deposits (arXiv:2510.14445). arXiv. <https://doi.org/10.48550/arXiv.2510.14445>

Rongier, G., & Peeters, L. (2025b). Towards geological inference with process-based and deep generative modeling, part 2: Inversion of fluvial deposits and latent-space disentanglement (arXiv:2510.17478). arXiv. <https://doi.org/10.48550/arXiv.2510.17478>

Geological Modeling from Legacy Seismic Data with Consideration of Uncertainties

Daniel Satizabal ¹, Ítalo Gomes Gonçalves ², David Nathan ¹, and Florian Wellmann ^{1,3}

¹ Chair of Computational Geoscience, Geothermics and Reservoir Geophysics, Georesources and Materials Engineering, RWTH Aachen University, Aachen, 52074, Germany; ² Geological Modeling Laboratory, Federal University of Pampa, Caçapava do Sul, RS, Brazil; ³ Fraunhofer Research Institution for Energy Infrastructures and Geothermal Systems (IEG), Bochum, 44801, Germany

Geological modeling is a fundamental component of reservoir characterization in geothermal exploration. A structural geological model, usually referred to as a geological model or geomodel, aims to represent the spatial relation between geological features such as rock unit boundaries, horizons, discontinuities (unconformities, faults) and intrusions at various scales [5]. In this contribution, we show how surrogate models can be employed to consider spatially correlated uncertainties in reprocessed legacy seismic data. Due to limited information in depth, geological models can contain significant uncertainties. In the KarboEx2-project, legacy seismic data from former coal exploration in the region of North Rhine Westphalia are digitized and reprocessed with modern seismic processing workflows. We investigate how uncertainties in the interpretation of this legacy data can be considered in subsequent geological modeling workflows.

In the context of model construction, this type of uncertainty relates to the real position of the input points, commonly associated with data uncertainties (seismic processing, picking and interpretation, etc.) [5]. Several position points result from the picking of the horizons on legacy seismic data. A simple way to address this type of uncertainty is to perform sampling from the data treating it as fully correlated (i.e., moving all points simultaneously) or fully uncorrelated (i.e., moving all points independently). However, geological errors are commonly correlated with distance.

One possibility to consider spatial correlations is to generate a geological surrogate model with a lower-dimensional representation of the modelled interface. In addition to accounting for different uncertainties in space, such a low-dimensional representation allows to perform inference, sensitivity analysis, etc. We explore here a workflow based on the application of a Variational Gaussian Process (VGP) model from inducing points using an open-source Python package (GeoML) [1], [2], [3], [4]. The Gaussian process model can provide uncertainty estimates and be trained on small datasets. However, it presents a major difficulty modeling complex data sets (as seismic interpretation) using stationary covariances [4]. Therefore, a way to address this limitation is the application of deep Gaussian process through non-stationary covariances, which better captures complex, spatially varying structural patterns.

Our results show that it is possible to create surrogate models efficiently for a range of geological settings – with a balance between the dimension (input points) of the surrogate model and the level of complexity of the original interface. In addition, due to a variational approach, uncertainties in the input data are naturally propagated into the surrogate model and thus into derived geological realizations. We outline how this AI driven, uncertainty aware modelling approach can be integrated into geothermal exploration workflows, including the uncertainties in the legacy seismic data.

1. Gonçalves, Í. G., Guadagnin, F., Kumaira, S., & Da Silva, S. L. (2021). A machine learning model for Structural Trend Fields. *Computers and Geosciences*, 149, 104715. <https://doi.org/10.1016/j.cageo.2021.104715>

2. Gonçalves, Í. G., Guadagnin, F., & Cordova, D. P. (2022). Learning spatial patterns with variational gaussian processes: Regression. *Computers and Geosciences*, 161, 105056. <https://doi.org/10.1016/j.cageo.2022.105056>

3. Gonçalves, Í. G., Guadagnin, F., & Cordova, D. P. (2023). Variational gaussian processes for implicit geological modeling. *Computers and Geosciences*, 174, 105323. <https://doi.org/10.1016/j.cageo.2023.105323>

4. Gonçalves, Í. G., and Wellmann, F. (2025). Uncertainty propagation in deep Gaussian Process Networks. *Mathematical Geosciences*, 57(6), 1115–1133. <https://doi.org/10.1007/s11004-025-10187-4>

5. Wellmann, F., and Caumon, G. (2018). 3-D Structural Geological Models: Concepts, methods, and uncertainties. *Advances in Geophysics*, 1–121. <https://doi.org/10.1016/bs.agph.2018.09.001>

A global delta map informed by machine learning

Nirubha Raghavi Thillaigovindarasu¹, Kim Cohen¹, Hans Middelkoop¹, Jaap Nienhuis¹

¹ Utrecht University

Deltas have long been hotspots of human civilisation due to its fertile lands, courtesy of riverine sediment that accumulate on the delta plain (Anthony et al. (2025)). Despite significant research interest in deltas, their global extent remains poorly mapped. Is the Netherlands a delta? And if so, where does it start and where does it end? Existing maps are often either localised and areally sporadic or are overly generalised (Tessler et al. (2015); Caldwell et al. (2019)). The objective of our study is to create a scientifically informed global dataset of delta plain areas. To that end, we applied a random forest classification based on global datasets of elevation, lithology, land cover, river networks, and features derived from such datasets. Training data for the classification model comes from a collection of existing delta maps. We apply our model along the world's coastal areas, forming probability maps of delta existence and extent, and to identify distinct deltas. We find relatively high F1 skills scores, and notable heterogeneity in the retrieved feature importance, with elevation, lithology, and fluvial proximity constituting the primary determinants of delta area. Our methodology thus enabled systematic delineation of delta plains worldwide, yielding a global dataset that integrates geomorphological principles with data-driven classification.

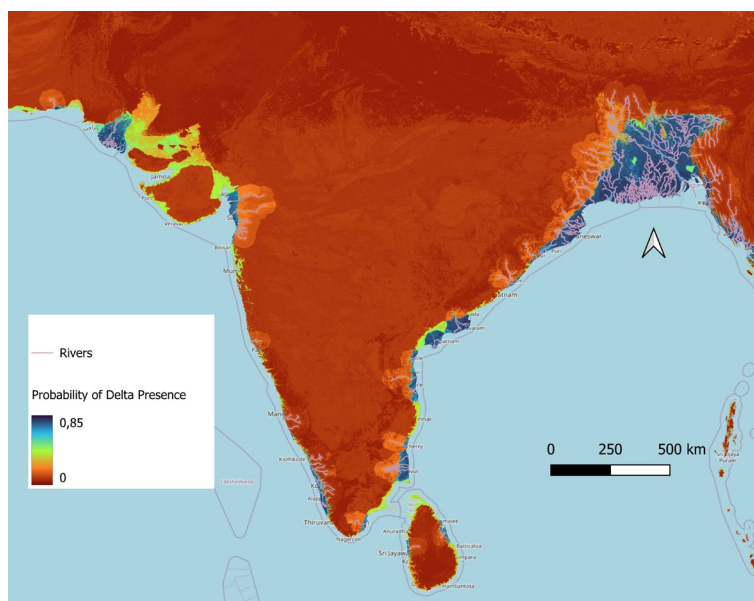


Figure 1: Probability map of delta presence in the Indian region

Anthony, E., Syvitski, J., Cohen, K. M., Saito, Y., Zăinescu, F., Vespremeanu-Stroe, A., ... & Chen, Z. (2025). A 7000-year record of human influence on Global River Deltas: Geomorphology, stratigraphy, the Anthropocene overprint and future. *Earth-Science Reviews*, 105302.

Caldwell, R. L., Edmonds, D. A., Baumgardner, S., Paola, C., Roy, S., & Nienhuis, J. H. (2019). A global delta dataset and the environmental variables that predict delta formation on marine coastlines. *Earth Surface Dynamics*, 7(3), 773-787.

Tessler, Z. D., Vörösmarty, C. J., Grossberg, M., Gladkova, I., Aizenman, H., Syvitski, J. P. M., & Foufoula-Georgiou, E. (2015). Profiling risk and sustainability in coastal deltas of the world. *Science*, 349(6248), 638-643.

AI Approaches for Geological Mapping: From Lithological Prediction to Geophysical Modelling

S. Tirdad¹, V. Silva dos Santos², E. Gosselin², E. Penda Biondokin², E. Gloaguen², G. Bellefleur³, M. Bavand Savadkoobi², F. Yrro³

1) Geological Survey of Canada (GSC)- Natural Resources Canada (NRCan), 490 Couronne St, Québec City, Quebec G1K 9A9, Canada

2) Institut National de la Recherche Scientifique, Eau Terre Environnement (INRS-ETE), 490 Couronne St, Québec City, Quebec G1K 9A9, Canada

3) Geological Survey of Canada (GSC)- Natural Resources Canada (NRCan), 601 Booth St., Ottawa, ON K1A 0G1, Canada

Geological mapping relies on the integration of diverse datasets, including remote sensing, geophysical surveys, and field observations. With recent advances in artificial intelligence (AI), there is growing potential to improve how these data are interpreted and translated into geological maps. However, many of these approaches remain at different stages of development and are often explored independently.

Here, we present a set of ongoing and completed research efforts that explore how AI can be applied to geological mapping and modelling.

We introduce a deep learning approach for predictive lithological mapping based on the Spatially Constrained Bayesian Network (SCB-Net) [1], which integrates remote sensing data with field observations to generate spatially consistent lithological maps while quantifying uncertainty. The method has been tested on datasets from northern Quebec and demonstrates improved accuracy compared to conventional architectures, while ensuring consistency with field data.

We also present ongoing work on deep-learning-based inversion of magnetic data [2], aiming to recover three-dimensional subsurface physical properties. This approach incorporates physics-informed constraints to reduce the non-uniqueness of inversion and improve geological interpretability.

Finally, we outline ongoing research in predictive modelling for mineral prospectivity [3], which integrates geological, geochemical, and geophysical datasets to estimate the spatial likelihood of mineralization and associated uncertainty.

These projects are developed as complementary efforts addressing key challenges in geological mapping: prediction of surface geology, reconstruction of subsurface structure, and identification of mineral potential. Together, they highlight both the potential and current limitations of AI in geological mapping, particularly in terms of data integration, uncertainty quantification, and transferability across regions.

[1] dos Santos, V. S., Gloaguen, E., & Tirdad, S. (2025). Lithological mapping using Spatially Constrained Bayesian Network (SCB-Net): A deep learning model for generating field-data-constrained predictions with uncertainty evaluation using remote sensing data. *Computers & Geosciences*, 204, 105964.

[2] Tirdad, S., Bellefleur, G., Yrro, F., Bavand Savadkoobi, M., and Gloaguen, E.: Toward Robust Three-Dimensional Magnetic and Gravity Inversion Using Deep Learning, EGU General Assembly 2026, Vienna, Austria, 3–8 May 2026, EGU26-1760, <https://doi.org/10.5194/egusphere-egu26-1760>, 2026.

[3] Gosselin, E., Gloaguen, E., & Tirdad, S. (2025). HANDLING MISSING GEOCHEMICAL DATA FOR PREDICTIVE MAPPING OF VMS DEPOSITS: A CASE STUDY OF THE BATHURST MINING CAMP (NB, CANADA). *Geoscience Canada*, 52(2).

A comparative study of machine learning approaches for lithostratigraphic interpretations of CPT data

Tom Van Haren^{*1}, Pieter Jan Kerstens², Filip Schouwenaars², Hossein Ghorbanfekr³, Katrijn Dirix¹, Prof. Dr. Dirk Valkenborg⁴

¹VITO [Environmental Intelligence, tom.vanharen@vito.be]

²VITO [Data & Analytics]

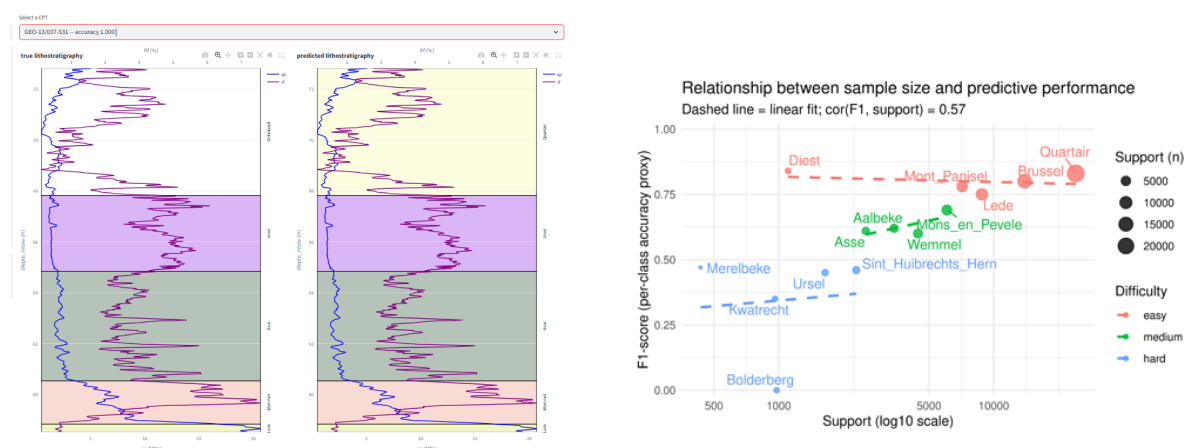
³IMEC

⁴Data Science Institute, UHasselt

Cone penetration tests (CPTs) provide quick and detailed information about the subsurface and are widely used in geotechnical engineering. They also help to provide detailed insight into lithostratigraphic correlation and interpretation for the development of urban subsurface models. However, lithostratigraphic interpretation of CPT data has remained largely a manual process, making it a time-consuming and subjective process. Automated and reproducible interpretation workflows are clearly needed to help put more detail into shallow urban subsurface models.

Our study explores the potential of machine learning models to support geologists and geo engineers with lithostratigraphic interpretation of CPT data on a large scale. It focuses on the development, comparison and integration of complementary approaches. A reference set of around 200 CPT's from the northern periphery of Brussels was manually interpreted by expert geologists. These CPT signals were subsequently enriched with a wide range of derived features and expert-informed constraints focusing on geological layer properties and sequences.

The presentation of this study will focus on the comparison and combination of different approaches with promising accuracies (>70%), demonstrating that this is an interesting way of working with supervised learning approaches to effectively capture lithostratigraphic patterns in CPT data. We will present different solution strategies encountered along the way and how they failed. We then present the latest results, with particular attention to model interpretability, and the role of expert feedback in assisted interpretation workflows.



Dashboard assisted CPT lithostratigraphy (left) and the effects of lithostratigraphy and sample size on predictive performance (right)

Extracting structured data from legacy borehole logs

Stijn Vermeeren

Swisstopo

In Switzerland, as in many other countries, a large quantity of geological data remains available only in unstructured form, such as scanned documents. This limitation is particularly relevant for borehole logs, as valuable data from thousands of wells is locked away in unstructured files that cannot be used for geological modelling or other data-centric applications. Entering such data into a structured digital database used to be a tedious manual job that was too slow and too costly to perform on a large scale. Thankfully, modern artificial intelligence methods, particularly those from the fields of computer vision and natural language processing, are opening new possibilities to increasingly automate the extraction of structured data from borehole logs and other geological documents.

Since 2023, the Swiss Federal Office of Topography swisstopo has gained significant experience in developing solutions for the automatic extraction of structured geological data. Some of these solutions are already used in production, enabling significant efficiency gains. All of our work is published as open-source software, enabling individuals and institutions around the world to run the automatic data extraction pipeline on their own data or even to contribute to the software's ongoing development.

In this talk, we will provide an overview of the geological data extraction solutions that have been developed so far at swisstopo. We will discuss the most successful methods and explain why chatbots, despite their huge successes in other fields, struggle with the correct interpretation of borehole logs. Finally, we will look at how generalizable our solution is, in terms of its ability to deal with boreholes logs with different templates, from different countries and in different languages.

Stratigraphie aus Profil extrahieren

Tiefe	Lithologische Beschreibung
0 m	Belag, sauberer Kies mit reichlich bis viel Sand
0.15 m	
0.15 m	brauner feinsandiger Silt, schwach tonig mit reichlich bis viel Kies; schwach humös; mit Ziegelresten (künstl. Auffüllung)
0.85 m	
0.85 m	dunkelbrauner stark tonig-siltiger Sand mit viel Kies und Steinen, Korn kantig, humos, verwittert, weich bis steif (Auffüllung und natürliche Deckschicht)
1.7 m	
1.7 m	grauer schwach siltiger Kies mit reichlich bis viel Sand, reichlich Steinen bis Ø 18; Korn kantengerundet bis gerundet, fein-mittel Kies vorherrschend, Blockhorizont an der Basis
4.2 m	
4.2 m	braun-grauer siltiger bis stark siltiger Kies mit reichlich bis viel Sand, wenig Steinen und vereinzelt Blöcken, Korn kantig bis gerundet - det z.T. mit Schluff, z.T. Korn verbacken, ürtliche Verkittung, Block bei 6.60 m Ø 30 cm
12 m	

Geotechnisches Institut AG **Rotationsbohrung RB3** **Moss**

Gartenstrasse 13 3007 Bern 031 / 25 57 33

Gt.Nr.: 188.2473 Objekt: Buehler AG, Wabern

Höhe in m ü M.: 560.57 Aufnahme: F. Keller

Koordinaten: 27250/495-170 Datum: 11./12.7.1988 Bohrmeister:

Tiefe	Profil	Labo Nr	w %	Geologie	Geotechnische Beschreibung
USCS	F t/m				
0.15	[Profil]			a	Belag, sauberer Kies mit reichlich bis viel Sand
0.85				b	brauner feinsandiger Silt, schwach tonig mit reichlich bis viel Kies; schwach humös; mit Ziegelresten (künstl. Auffüllung)
1.70					dunkelbrauner stark tonig-siltiger Sand mit viel Kies und Steinen, Korn kantig, humos, verwittert, weich bis steif (Auffüllung und natürliche Deckschicht)
4.20		3			grauer schwach siltiger Kies mit reichlich bis viel Sand, reichlich Steinen bis Ø 18; Korn kantengerundet bis gerundet, fein-mittel Kies vorherrschend, Blockhorizont an der Basis
12.00		4		c	braun-grauer siltiger bis stark siltiger Kies mit reichlich bis viel Sand, wenig Steinen und vereinzelt Blöcken, Korn kantig bis gerundet, det z.T. mit Schluff, z.T. Korn verbacken, ürtliche Verkittung, Block bei 6.60 m Ø 30 cm

Bemerkung:
Bohrloch trocken

1 / 1

Abbrechen Stratigraphie hinzufügen

Imaging Subsurface Reservoirs for Hydrogen, Geothermal, and Critical Minerals: Multi-Physics Imaging and AI-Assisted Interpretation

Velimir (Monty) Vesselinov, Tracy Kliphuis

EnviTrace LLC, Santa Fe, New Mexico, USA

The accelerating demand for low-carbon energy and critical mineral resources is driving renewed interest in subsurface systems capable of hosting geothermal fluids, geologic hydrogen accumulations, and mineralized reservoirs. These systems are controlled by complex geologic structures, fracture networks, and fluid pathways that are difficult to characterize using single-method approaches. Advances in multi-physics geophysical imaging, combined with AI-assisted interpretation, are enabling a new generation of integrated subsurface characterization workflows. Our work examines how the integration of seismic, electromagnetic, gravity, and well data can be used to image reservoir architecture across multiple spatial scales. By jointly analyzing complementary geophysical signals, multi-physics approaches help resolve structural controls, permeability pathways, thermal regimes, and mineralization patterns that govern the formation and productivity of subsurface reservoirs. Recent developments in machine learning and physics-informed AI provide powerful tools for interpreting these complex datasets. AI-assisted workflows can identify subtle patterns, integrate heterogeneous datasets, and accelerate inversion and interpretation processes while remaining constrained by geologic and physical principles. These approaches improve the detection of fracture systems, fluid migration pathways, and alteration zones that are critical for hydrogen systems, geothermal reservoirs, and critical mineral exploration. Case studies demonstrate how multi-physics imaging combined with AI-driven interpretation can reduce exploration uncertainty and improve the prediction of subsurface resource potential. The integration of geophysical sensing, computational methods, and geologic knowledge provides a unified framework for imaging and evaluating energy and mineral systems in the subsurface.

