https://doi.org/10.5194/egusphere-2025-1294
Preprint. Discussion started: 26 May 2025

Recovery of stratigraphic data with associated uncertainties from drillhole databases using litho2strat 1.0

Vitaliy Ogarko[1,2] and Mark Jessell[1,2]

[1]Centre for Exploration Targeting (School of Earth Sciences), The University of Western Australia, Crawley, 6009 WA, Australia

[2]Mineral Exploration Cooperative Research Centre, The University of Western Australia, Crawley, 6009 WA, Australia

Correspondence: Vitaliy Ogarko (vitaliy.ogarko@uwa.edu.au)

https://doi.org/10.5194/egusphere-2025-1294
Preprint. Discussion started: 26 May 2025

Abstract

Australian commonwealth, state and territory geological surveys possess information on over 3 million drillhole logs. There are many more wells drilled in the search for oil and shallower holes related to hydrogeology. Other countries no doubt have similar data holdings. Together these legacy drillhole datasets have the potential to significantly improve our subsurface data coverage but have limited use as constraints on regional 3D geological models as many if not most drill logs lack stratigraphic information.

This study develops open-source codes and methodologies for stratigraphy recovery from drillhole databases by introducing a correlation algorithm that integrates data from multiple drillholes. The algorithms combine constraints from lithological descriptions, with stratigraphic relationships automatically derived from regional maps. In addition, by integrating uncertainty quantification and presenting multiple geological hypotheses, the resulting stratigraphical description provide critical insights for resource estimation, scenario analysis, and data acquisition strategies.

The application of our method to a dataset of 52 drillholes from South Australia demonstrated its ability to make useful predictions of stratigraphic solutions and quantifying associated uncertainties. These results not only validate our approach but also highlight opportunities to refine current stratigraphic descriptions and provide a valuable new source for regional 3D geological modelling.

https://doi.org/10.5194/egusphere-2025-1294
Preprint. Discussion started: 26 May 2025

1. Introduction

Drillhole data serves as a fundamental constraint for subsurface geological exploration and 3D geological modelling, offering direct insights into lithological and hence stratigraphic features (Wellmann & Caumon, 2018). However, the inherent sparsity of such data, coupled with challenges posed by legacy datasets maintained by industry and Geological Survey Organizations (GSOs), often hinders comprehensive geological understanding and modelling (Jessell et al., 2010; Pakyuz-Charrier et al., 2018). GSOs' databases typically contain complexly coded lithological information but limited stratigraphic data, creating a critical gap in the data needed for accurate and meaningful geological predictions (Hartmann & Moosdorf, 2012).

Geological modelling plays a crucial role in understanding subsurface structures and processes, providing a foundation for various applications in earth sciences (M. Jessell et al., 2014). Such modelling commonly relies on datasets such as borehole data, geophysical data, and mapping data. From these, borehole data provide the most accurate insights into subsurface geology and stratigraphy (Guo et al., 2022). The models generated through geological modelling can serve dual purposes: they can be directly employed for geological interpretations, such as identifying fault systems, and mineral deposits (Alvarado-Neves et al., 2024; Vollgger et al., 2015), or they can be integrated as constraints in methodologies that use a prior 3D model, such as geophysical inversions (Giraud et al., 2017; Martin et al., 2024; Ogarko et al., 2021; Tarantola, 2005) and hydrogeological forward modelling (D'Affonseca et al., 2020). By incorporating geological models into geophysical inversion, it is possible to refine the interpretation of subsurface properties and achieve greater accuracy in representing complex geological environments.

Modern drillhole measurement techniques primarily focus on chemical, mineralogical and lithological characterization, whereas the fundamental categorical unit of regional 3D geological models is defined by its stratigraphy (Calcagno et al., 2008; Caumon et al., 2009; Mallet, 2002). This discrepancy underscores the need for innovative approaches to recover and integrate stratigraphic information from existing datasets. Recent advancements in automation, particularly through machine learning and natural language processing, have demonstrated significant potential in addressing these challenges by standardizing and extracting lithological and stratigraphic data at scale (Guo et al., 2024; Joshi et al., 2021; Schetselaar & Lemieux, 2012).

This study develops open-source codes and methodologies for stratigraphy recovery from drillhole databases by introducing a correlation algorithm that integrates data from multiple drillholes, we enhanced the robustness and reliability of stratigraphic interpretations. By integrating uncertainty quantification and presenting multiple geological hypotheses, the resulting stratigraphical description provide critical insights for resource estimation, scenario analysis, and data acquisition strategies. The application of our method to a dataset of 52 drillholes from South Australia demonstrated its ability to make useful predictions of stratigraphic solutions and quantifying associated uncertainties. These results not only validate our approach but also highlight opportunities to refine current stratigraphic descriptions, improving workflow accessibility and paving the way for more effective geological assessments and decision-making processes.

https://doi.org/10.5194/egusphere-2025-1294
Preprint. Discussion started: 26 May 2025

2. Methodology

2.1 Workflow

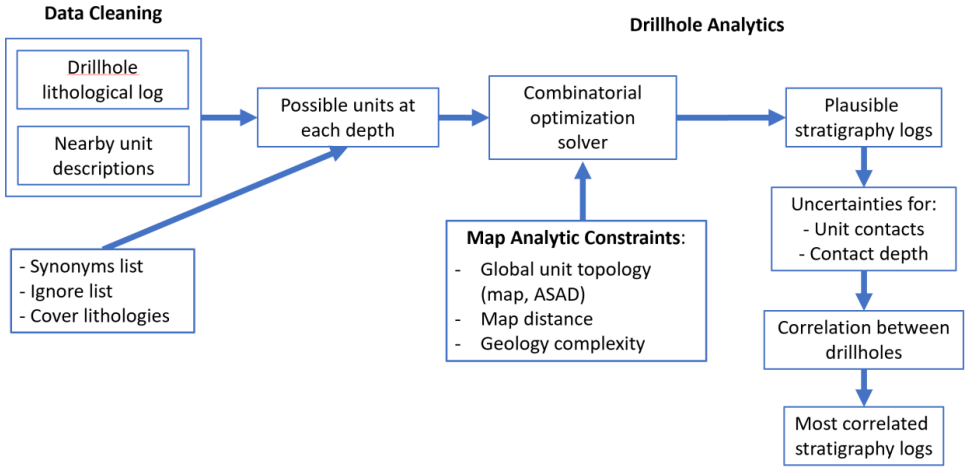


Figure 1: The different stages of the analysis.

The workflow shown in Fig. 1 consists of 5 key steps grouped into three main tasks: Data Cleaning (using the dh2loop code), Map Analytic Constraints (using map2loop and custom codes developed for this project) and Drillhole Analytics (using the litho2strat code developed for this project).

2.1.1 Data Cleaning

Prior to analysing the drillhole data we went through a number of automated data cleaning and harmonisation steps.

a) Harmonisation of drillhole lithology descriptions using the dh2loop code described in (Joshi et al., 2021) (code available here: https://github.com/Loop3D/dh2loop) This enables us to produce a standardised lithological description for multiple drillholes in a region, regardless of their provenance. This includes the use of a synonym list ("granite" vs "granitoid"), and ignore list (e.g. "fault") together with a list of cover lithology terms (e.g. "saprolite") that enables us to simplify the list of terms and exclude irrelevant information.

b) Harmonisation of lithological descriptions for formations described in the geological map of the target area. This ensures that the same terminology is used for borehole lithological descriptions and map lithologies.

Together steps a & b provide a list of possible units at each depth down a drill hole.

2.1.2 Map Analytic Constraints

a) Calculation of the distance between each polygon in a map and the target borehole. A custom Python script was developed. This information can be used as a guide to the likelihood that a drillhole would intersect a given unit.

https://doi.org/10.5194/egusphere-2025-1294
Preprint. Discussion started: 26 May 2025

b) We then used the map2loop engine (M. Jessell et al., 2021) (code available here: https://github.com/Loop3D/map2loop) to extract the topological relationships between the surface expression of stratigraphic different units. This would later be used to assess the likelihood that two units would be in contact in the drillhole.

Unit connectivity information can also be obtained from the Australian Stratigraphic Units Database (ASUD) as well as from various published reports containing stratigraphic data. The ASUD serves as a comprehensive repository of geological information, providing valuable insights into the relationships between different stratigraphic units across Australia. Additionally, numerous geological surveys and research studies offer stratigraphic data that can further enrich our understanding of unit connectivity. By leveraging this information, we can enhance our stratigraphic models, improve the accuracy of correlations between drillholes, and facilitate a deeper understanding of the geological framework in specific regions.

These two steps allow us to capture information on the spatial and topological relationships between the mapped units.

2.1.3 Drillhole Analytics

a) In this stage of the analysis, we developed a new code called litho2strat (code available here: https://github.com/Loop3D/litho2strat; Ogarko et al., 2025). The information about code design is detailed in Section 2.4. The code uses a combinatorial optimisation solver as follows:

- We employed a Branch and Bound algorithm, which is a powerful method for solving optimization problems (Land & Doig, 1960). This approach systematically enumerates all potential candidate solutions by exploring the state space.
- New branches could be added while traversing the drillhole log from top to bottom, using a list of possible units at each depth down a drill hole and a unit topological relationship.
- We applied bounds by adding solution constraints. The branch is discarded if it does not satisfy the constraints. The types of constrains we apply are discussed in Section 2.2.
- In the end, we could obtain a plausible set of solutions fitting the data and satisfying the constraints. This strategy not only ensures a thorough exploration of the solution space but also optimizes efficiency by eliminating unnecessary computations, ultimately leading to a more effective and streamlined search process.
- Finally, we calculated the uncertainties directly from the whole set of solutions.

b) In this stage, we employ the litho2strat code to establish correlations between drillholes, which helps to reduce uncertainty in the calculated set of plausible stratigraphic solutions. The correlated solution scores are assigned based on overall uncertainty, with solutions receiving the highest scores selected as the most plausible options. For further details on the correlation algorithm utilized, please refer to Section 2.3.

https://doi.org/10.5194/egusphere-2025-1294
Preprint. Discussion started: 26 May 2025

2.2 Solution constraints

For the Branch and Bound (BnB) algorithm, providing efficient constraints (bounds) is crucial for generating geologically plausible stratigraphies and reducing the search space. Without these bounds, the BnB algorithm reverts to exhaustive search, which is less efficient. We utilize two types of solution constraints: the first can be derived from geological maps (as discussed in the 'Map Analytic Constraints' section), while the second is selected by the user based on the expected geological complexity of the area (e.g., the presence of known faults).

The specific constraints employed include:

1. **Distance from the Drillhole to the Unit**: This constraint limits the number of geological units considered in the calculations, ensuring relevance to the drillhole's location.
2. **Global Unit Connectivity**: This constraint restricts the potential contacts between units, enhancing the accuracy of the stratigraphic model.
3. **Top Unit Constraint**: Information regarding the top unit can be extracted from geological maps, providing a foundational boundary for the stratigraphy.
4. **Solution Complexity (Occurrence)**: This constraint sets a maximum limit on how many times a unit can appear in the final stratigraphy. For example, in the presence of a fault, the same geological unit may appear twice within the same log.
5. **Solution Complexity (Contacts)**: This constraint defines the maximum number of unit contacts allowed within a continuous sequence of drillhole lithologies. For instance, in a continuous sequence of "sandstone" lithologies, different sandstones may belong to separate geological units.

These constraints work together to enhance the efficiency and effectiveness of the BnB algorithm, ensuring that the resulting stratigraphies are both plausible and reflective of the geological context.

2.3 Solution correlation

We utilize solution correlation analysis to establish relationships between multiple drillholes, serving as a constraint on the plausibility of stratigraphic ordering. This correlation can be applied either to stratigraphy logs or to the topological relationships of units, represented through connectivity graphs.

A key challenge in correlating stratigraphy logs is that units at the same depth may not align across different drillholes, mainly due to variations in unit dip and thickness. To address this, we focus on correlation based on topological relationships. Connectivity graphs are constructed for each drillhole using a plausible set of stratigraphic solutions derived from the Branch and Bound (BnB) solver.

In these graphs, the edges are weighted by the probability of unit contacts, calculated as the number of solutions that include a specific unit contact divided by the total number of solutions for that drillhole. This approach enhances the accuracy of the correlations, providing a more reliable framework for establishing stratigraphic relationships.

https://doi.org/10.5194/egusphere-2025-1294
Preprint. Discussion started: 26 May 2025

We propose the correlation algorithm as follows:

I. **Calculate Solution Scores**: For each solution i, compute the solution score $S(i, G_{self})$ using the connectivity graph from the current drillhole G_{self}. This score is defined as the sum of the edge weights for all solution contacts, indicating the likelihood of the solution within the current drillhole.
II. **Calculate External Solution Scores**: For each external drillhole j, calculate the solution scores $S(i, G_{ext}^{j})$ using their respective connectivity graphs. This score reflects the likelihood of the solution for each external drillhole.
III. **Build Final Correlated Solution Score**: Construct the final correlated solution score $C(i)$ using the following formula:

$$C(i) = \frac{S(i, G_{self}) + \sum_{j=1}^{N-1} S(i, G_{ext}^{j})}{N}$$

This final score represents the overall likelihood of the solution across all drillholes, effectively correlating the current solution with all solutions from the external drillholes based on the topological relationships of the geological units.

The final correlated solution score provides a means to assign correlated scores to various solutions, effectively altering their ranking within the overall assessment. This adjustment not only enhances the clarity of which solutions are more viable but also helps to reduce uncertainty associated with those solutions. By refining the evaluation process in this manner, we can achieve a more precise understanding of the relationships between different geological units. By integrating and correlating drillhole data effectively, we ensure that the stratigraphic framework accurately reflects the natural variations and interconnections present in the subsurface. This consistency is essential for making informed decisions in geological exploration ultimately leading to more reliable outcomes in our geological interpretations.

In the proposed correlation algorithm, we can select between summation and multiplication for calculating the correlated score in step iii. Summation is less sensitive to outliers; for instance, if one drillhole in the set has a zero score, using multiplication would result in a total score of zero. Additionally, we can apply weighting to different drillholes, such as by using geological distance, to reduce the influence of more distant drillholes on the overall score.

Moreover, it's important to note that the algorithm exhibits linear scalability with respect to the number of drillholes. This efficiency is achieved by utilizing solution connectivity graphs rather than evaluating each pair of solutions individually, which can be computationally expensive.

2.4 Code design

A Python package called litho2strat has been developed for stratigraphy recovery. It can be easily installed using the command "pip install", and it has minimal external library dependencies: numpy, matplotlib, and NetworkX. The NetworkX library is utilized to create a directed graph data structure that represents the topological relationships of relative unit ages (Hagberg et al., 2008). It also supports exporting graphs to GML format (Himsolt, 1997) for advanced graph visualization with tools like yEd (https://www.yworks.com/products/yed).

https://doi.org/10.5194/egusphere-2025-1294
Preprint. Discussion started: 26 May 2025

Interaction with the code is facilitated through a *Parfile*, a text file that contains all necessary parameters and paths to the input data files. The parameters in the *Parfile* are organized into several categories based on their functionality, including input file paths, solver settings, and data preprocessing options. An example of such a *Parfile* is provided in Appendix A.

The code architecture efficiently organizes distinct modules, including data reader, the user interface (represented by the *Parfile*), the algorithms (such as the solver), and the visualization components (e.g., output figures and graphs), as shown in Fig. 2. This design enhances code readability, making it easier for developers to understand and navigate the codebase. Additionally, it facilitates further extensions by allowing new features to be integrated seamlessly. The structure also supports effective testing, ensuring that modifications can be verified without introducing errors.

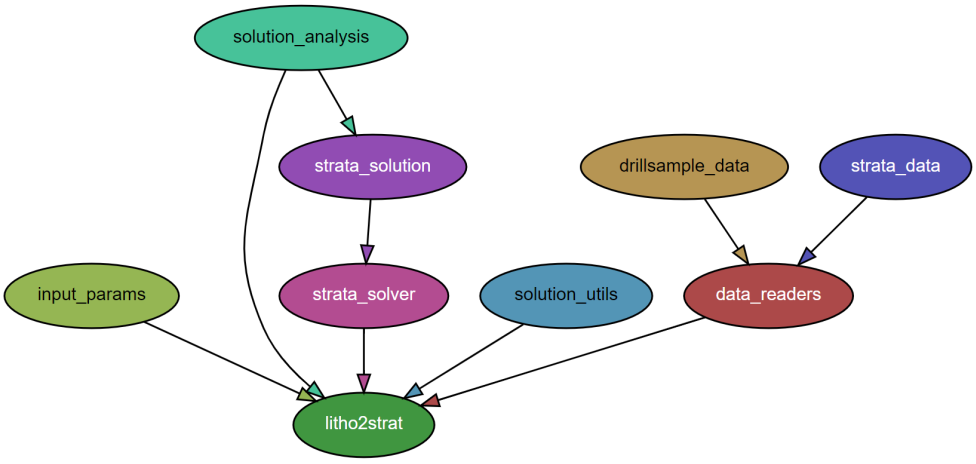


Figure 2: The module dependencies of the *litho2strat* code. The graph is generated by the *pydeps* utility, while excluding external dependencies.

3. Example Use

For this example, we used a set of 52 drillholes from South Australia originally drilled by Teck Cominco Pty. Ltd. (Fig. 3). This area was chosen as there were a number of holes equally spaced with a relatively homogenous spatial distribution and the holes provided both lithological logs and existing interpretations of the down-hole stratigraphy.

https://doi.org/10.5194/egusphere-2025-1294
Preprint. Discussion started: 26 May 2025

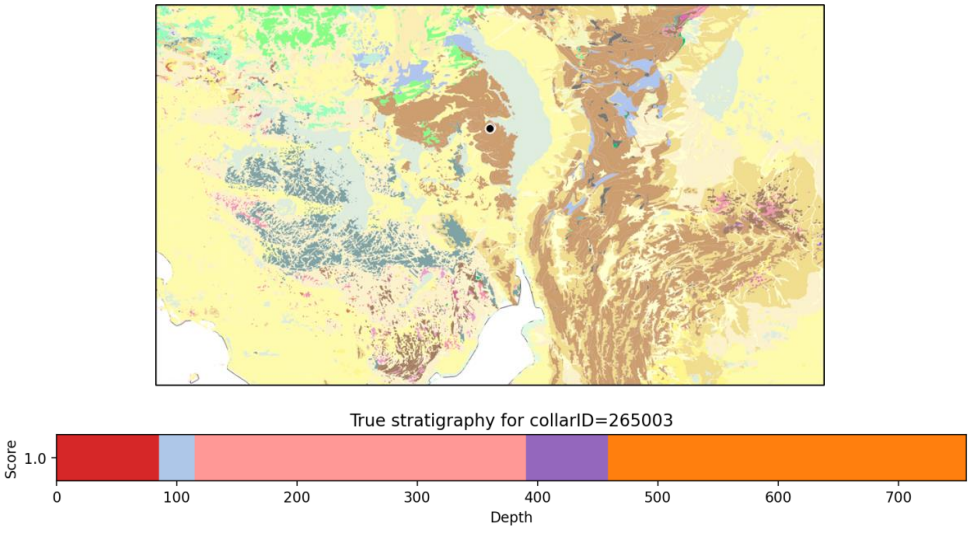


Figure 3: Location of South Australia test area, together with an example stratigraphic log.

Data Cleaning

Examples of terms in the ignore list for this case study include:

```
Breccia (Undiff. Origin)
Ironstone (Metasomatic)
No Information
Solution-Collapse Breccia
Vein (Undifferentiated)
```

Examples of the thesaurus of synonyms for this case study area include:

```
dolomite, dolomite rock, carbonate rock, limestone
conglomerate, diamictite
grit, sandstone, quartzite, siltstone
gabbro, gabbronorite
```

Map Analytics

The figure below shows stratigraphic units coloured as a function of the distance to one of the drillholes (Fig. 4). A large search area was used for this example as the stratigraphy is fairly flat lying so there is no guarantee that a unit will reach the surface in the local neighbourhood.

https://doi.org/10.5194/egusphere-2025-1294
Preprint. Discussion started: 26 May 2025

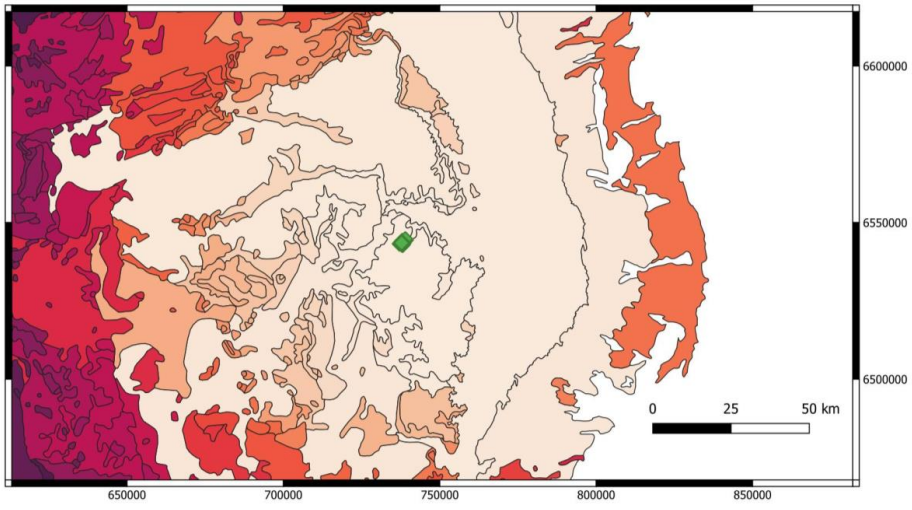


Figure 4. Distant of stratigraphic units to drillholes (darker signifies larger distance). Green diamonds show the location of the drillholes.

In the initial analysis we plotted a network graph of all know topological relationships between stratigraphic units based on the geology map (extending out 100 km from the test area), the ASUD database, and additional information from published reports (Fig. 5).

https://doi.org/10.5194/egusphere-2025-1294
Preprint. Discussion started: 26 May 2025

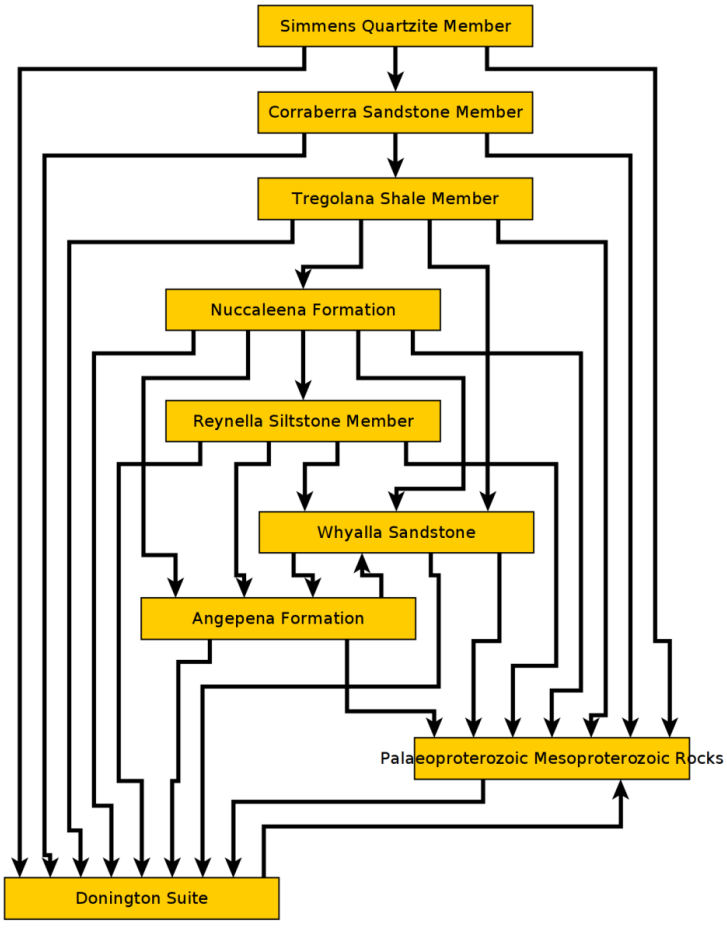


Figure 5: Topological relationships between units in and around the test area.

Drillhole Analytics

The drillholes analysis calculated every possible stratigraphic ordering that was consistent with the observed lithological ordering down the drillhole and solution constraints (described in Sec. 2.3). By collating the results for all possible solution paths, we can produce estimates of the probability that any depth interval will be a particular stratigraphic unit (Fig. 6).

https://doi.org/10.5194/egusphere-2025-1294
Preprint. Discussion started: 26 May 2025

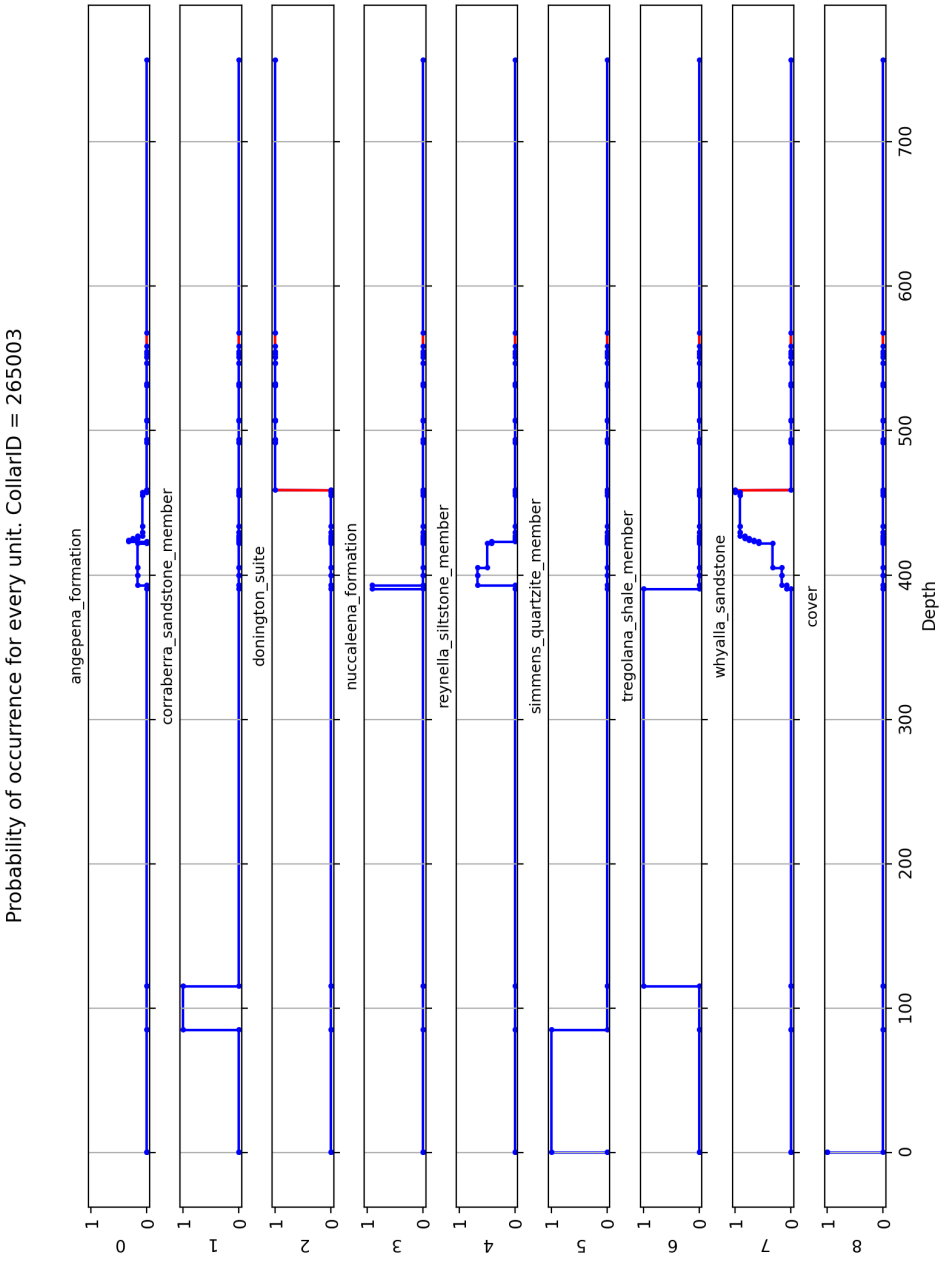


Figure 6: Estimated probability of each stratigraphic unit occurring at a given depth for a single drillhole.

In Fig. 7, we present the final (local) unit connectivity derived from the stratigraphy solutions generated. The width of the graph edges indicates the probability of unit contacts, with thicker edges

https://doi.org/10.5194/egusphere-2025-1294
Preprint. Discussion started: 26 May 2025

signifying higher probabilities. This visual representation allows for a clear comparison of connectivity before (Fig. 6) and after the stratigraphic analysis.

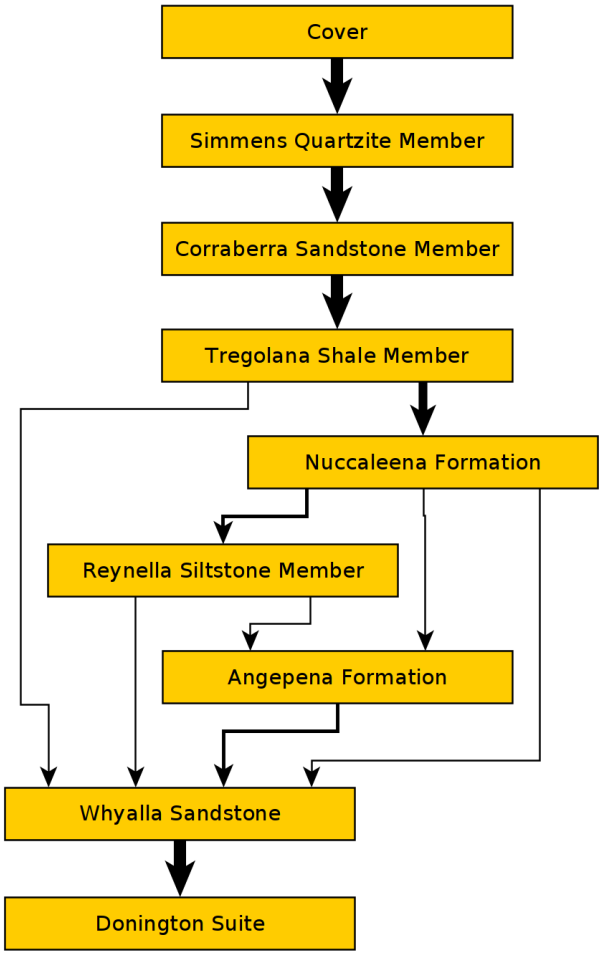
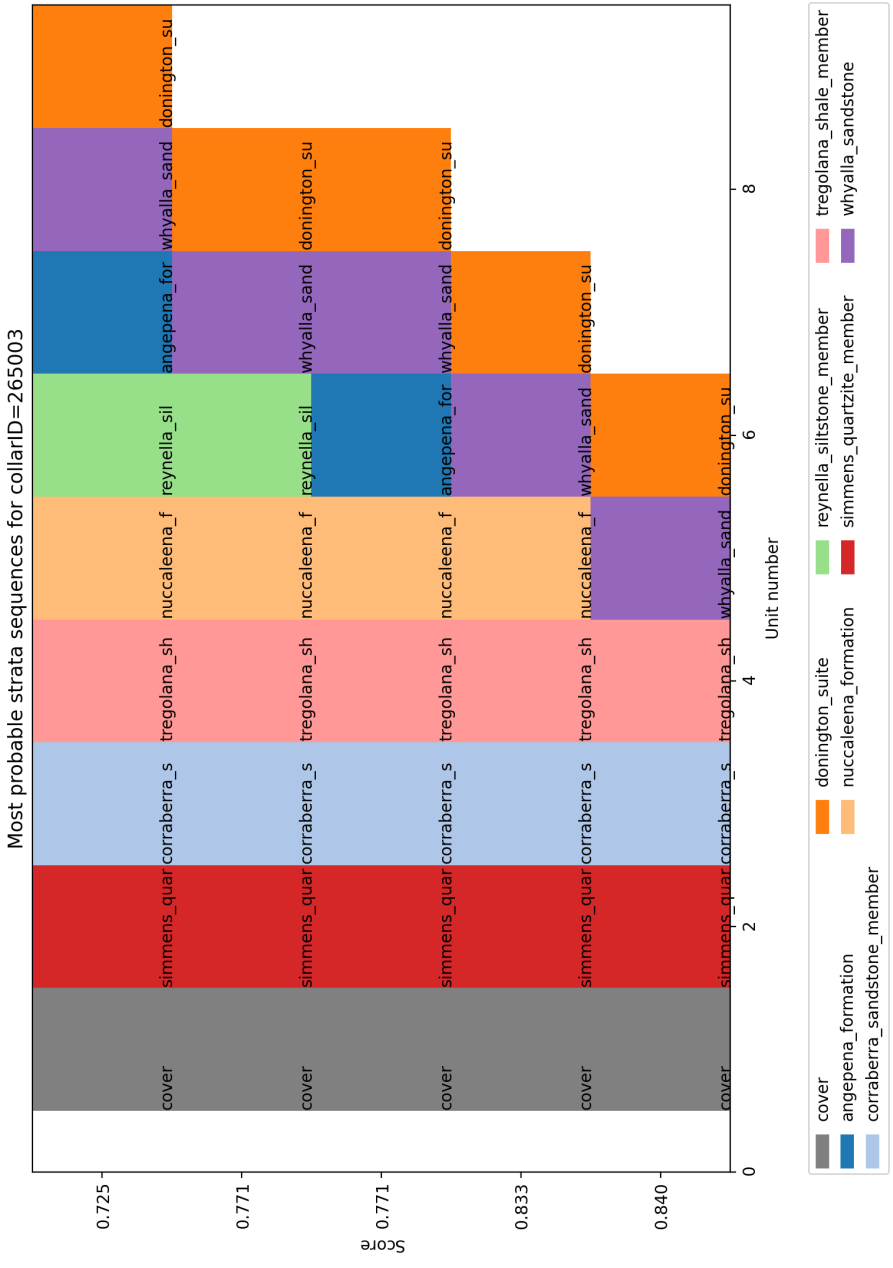


Figure 7: Calculated local topology using all solutions. Graph edges (relationships) between two stratigraphic units are displayed as a probability of a that contact-relationship occurring.

The final solution score for a single ordering is calculated by summing of the probabilities of the contact edge weights. This allows us to sort the orderings by probability, ignoring stratigraphic thickness for now (Fig. 8).

https://doi.org/10.5194/egusphere-2025-1294
Preprint. Discussion started: 26 May 2025



281

282 Figure 8: The 5 most probable stratigraphic orderings, with their solution probability on the x axis
283 and order of depth on the y axis.

284

https://doi.org/10.5194/egusphere-2025-1294
Preprint. Discussion started: 26 May 2025

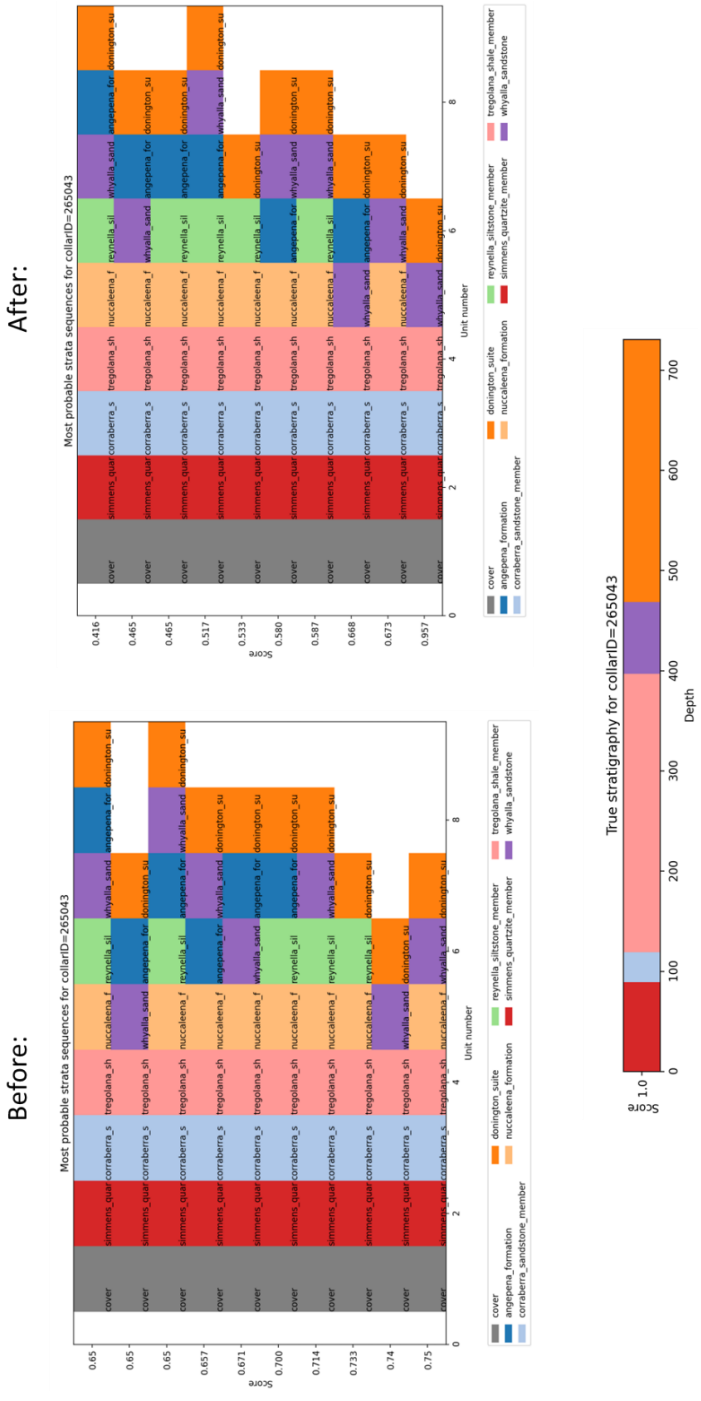
285 Finally, we can then include the depths to contacts between units in the drillhole based on the
286 previous analyses (Fig. 9).

287

288 Figure 9: The 12 most probable stratigraphic orderings showing true depth of contact (above)
289 compared to the stratigraphy as logged for the same hole.

290

https://doi.org/10.5194/egusphere-2025-1294
Preprint. Discussion started: 26 May 2025



291

292 Figure 10: Comparison of ordering for one hole (left) vs ordering for that hole considering the
293 outcomes of 45 other drillholes in the neighbourhood.

https://doi.org/10.5194/egusphere-2025-1294
Preprint. Discussion started: 26 May 2025

In the next stage of our analysis, we perform solution correlation across multiple drill holes to establish a plausible stratigraphic order and reduce uncertainty. Figure 10 illustrates the comparison of the most probable stratigraphies before and after correlation. Prior to correlation, the solution that aligns with the "true" stratigraphy (the correct solution) is ranked second, with a score of S=0.74, while the highest-ranked solution has a score of S=0.75. However, after applying the correlation, the correct solution rises to the top rank with a score of S=0.95, whereas the previously highest-ranked solution falls to second place with a score of S=0.67. This correlation analysis not only helped identify the correct solution but also significantly reduced its relative uncertainty, increasing the relative score between the top two solutions from 1% to 42%.

4. Discussion and Future Work

Whilst we were able to develop a workflow that successfully provided useful stratigraphic analyses for our test area, we recognise that for other areas the methodology was not always as successful. We have identified several aspects of the current stratigraphic descriptions that we think will significantly expand the useability of the workflow we present above.

1) Lithological Uncertainty. The principal reason for this was that the lithological descriptions of stratigraphies in many areas is quite vague. Successive stratigraphic units in a group might have very similar lithological descriptions.

   As an example, we look at the Hamersley Group, in Western Australia (Maldonado & Mercer, 2018). If we examine the GSWA explanatory notes for three successive formations (Mt McRae Shale, Mt Sylvia Formation and the Wittenoom Formation) in the GSWA explanatory notes their lithologies are described as:

   > **Mt McRae Shale** - Mudstone, siltstone, chert, iron-formation, and dolomite. Thin bands of shard-bearing volcanic ash in upper parts.
   >
   > **Mt Sylvia Formation** - Mudstone, siltstone, chert, iron-formation, and dolomite.
   >
   > **Wittenoom Formation** - Thinly bedded dolomite and dolomitic shale, with minor black chert, shale, banded iron formation and sandstone.

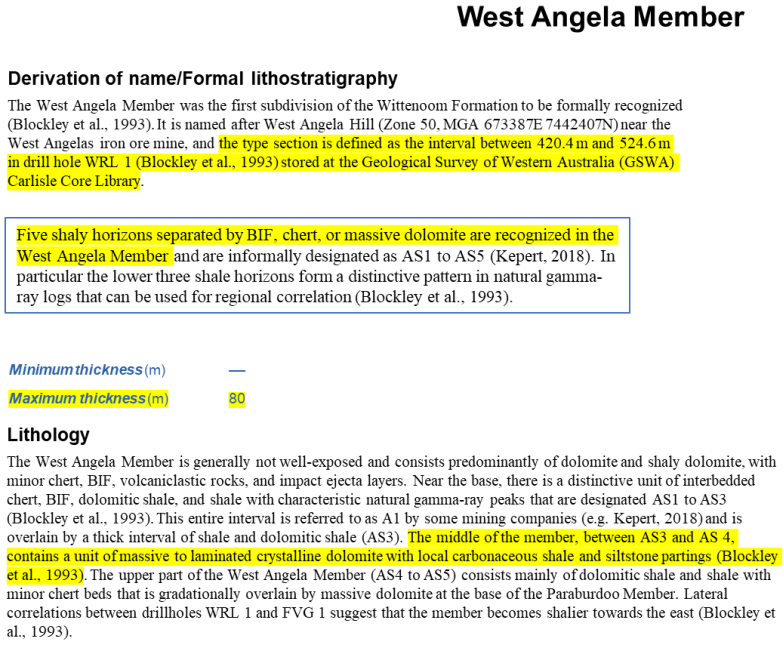
   We can see that there is a significant overlap in lithologies, with an ordering of lithologies but without constraints on the percentage of each lithology in the three formations. This additional information, even as an estimate, would provide useful constraints on the likelihood that a specific lithology is associated with a given stratigraphic unit.

2) Min-Max thickness estimates. In some areas, there is useful information on the minimum, maximum and average stratigraphic thickness of units.

3) Stratigraphic ordering of lithologies. Additional information on commonly occurring orderings of lithologies within a given formation or member would also provide useful constraints.

https://doi.org/10.5194/egusphere-2025-1294
Preprint. Discussion started: 26 May 2025

334



West Angela Member

Derivation of name/Formal lithostratigraphy

The West Angela Member was the first subdivision of the Wittenoom Formation to be formally recognized (Blockley et al., 1993). It is named after West Angela Hill (Zone 50, MGA 673387E 7442407N) near the West Angelas iron ore mine, and the type section is defined as the interval between 420.4 m and 524.6 m in drill hole WRL 1 (Blockley et al., 1993) stored at the Geological Survey of Western Australia (GSWA) Carlisle Core Library.

Five shaly horizons separated by BIF, chert, or massive dolomite are recognized in the West Angela Member and are informally designated as AS1 to AS5 (Kepert, 2018). In particular the lower three shale horizons form a distinctive pattern in natural gamma-ray logs that can be used for regional correlation (Blockley et al., 1993).

Minimum thickness (m) —
Maximum thickness (m) 80

Lithology

The West Angela Member is generally not well-exposed and consists predominantly of dolomite and shaly dolomite, with minor chert, BIF, volcaniclastic rocks, and impact ejecta layers. Near the base, there is a distinctive unit of interbedded chert, BIF, dolomitic shale, and shale with characteristic natural gamma-ray peaks that are designated AS1 to AS3 (Blockley et al., 1993). This entire interval is referred to as A1 by some mining companies (e.g. Kepert, 2018) and is overlain by a thick interval of shale and dolomitic shale (AS3). The middle of the member, between AS3 and AS 4, contains a unit of massive to laminated crystalline dolomite with local carbonaceous shale and siltstone partings (Blockley et al., 1993). The upper part of the West Angela Member (AS4 to AS5) consists mainly of dolomitic shale and shale with minor chert beds that is gradationally overlain by massive dolomite at the base of the Paraburdoo Member. Lateral correlations between drillholes WRL 1 and FVG 1 suggest that the member becomes shalier towards the east (Blockley et al., 1993).

335

336 Figure 11: Free-text descriptions of the West Angela Member in the GSWA Explanatory Notes.

337 All three of these types of information are often included in the free-text portions of stratigraphic
338 databases, such as the example shown for the West Angela Member in the GSWA Explanatory
339 Notes in Fig. 11. In this example the free-text provides more specific information on the thickness,
340 the ordering of lithologies and the relative proportions of lithologies. With the advent of more
341 sophisticated Machine Learning methodologies, the extraction of this ancillary data in a
342 standardised form from reports and the stratigraphic databases themselves will open up new
343 possibilities for constraining stratigraphy. Similarly, the codes developed in dh2loop for
344 harmonising lithological terminologies will expand greatly in coming years.

345 4) Inclusion of discontinuity information in the litho2strat workflow (most often logged faults)
346 could help to define where breaks in stratigraphy are most likely to occur
347
348 5) Inclusion of secondary descriptive information (for example grain size) could help to refine our
349 younging estimators in areas of uncertain facing.
350
351 6) There is no doubt that the advent of Large Language Models will have a profound effect on
352 our ability to extract and categorize information from unstructured data sources, and
353 algorithms based on these approaches will probably replace the data extraction and data
354 harmonisation modules in future versions of this workflow.
355
356
357

https://doi.org/10.5194/egusphere-2025-1294
Preprint. Discussion started: 26 May 2025

5. Conclusions

We developed codes and methodologies for stratigraphy recovery from drillhole databases, utilizing the branch and bound algorithm as a foundational framework. To ensure the generation of geologically plausible solutions, we implemented various types of constraints that account for the complexities of subsurface geology.

To further reduce uncertainty in the obtained solutions, we introduced a correlation algorithm that leverages information from multiple drillholes simultaneously. This innovative approach allows for a more robust analysis by integrating data across different locations, enhancing the reliability of the stratigraphic interpretations.

Our proposed method was applied to a dataset comprising 52 drillholes from South Australia. The results demonstrated that the algorithm successfully predicts the correct stratigraphic solution while providing associated uncertainty metrics, effectively validating its performance against measured stratigraphy data.

Additionally, we identified several key aspects of the current stratigraphic descriptions that could significantly enhance the usability of the workflow we have presented. These enhancements aim to improve the accessibility and applicability of our methodology, paving the way for more effective geological assessments and decision-making processes in the field.

https://doi.org/10.5194/egusphere-2025-1294
Preprint. Discussion started: 26 May 2025

Code and data availability. The software and datasets used in this study are publicly available for download at GitHub (https://github.com/Loop3D/litho2strat) and Zenodo (https://doi.org/10.5281/zenodo.15064469, Ogarko et al., 2025).

Author contribution. VO and MJ are the primary contributors to this study. VO led the research, developed the methodology and software, and prepared the manuscript. MJ provided guidance on drillhole data analysis and contributed to manuscript writing.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The work has been supported by the Mineral Exploration Cooperative Research Centre whose activities are funded by the Australian Government's Cooperative Research Centre Program. This is MinEx CRC Document 20**/***.

https://doi.org/10.5194/egusphere-2025-1294
Preprint. Discussion started: 26 May 2025

Appendix A- Control file for litho2strat code

Example usage: `python3 litho2strat.py -p ./parfiles/Parfile_SA.txt`

Example parfile:

```
[FilePaths]
topology_filename = data/SA_test_data/newpairs_20_06_2023.gml
cover_unit_filename = 
ignore_list_filename = data/SA_test_data/ignore_list.txt
alternative_rock_names_filename = data/SA_test_data/alternative_rock_names.txt
unit_colors_filename = data/SA_test_data/unit_colors.csv

drillsample_filename = data/SA_test_data/litho_tables/litho_$collarID$.csv
stratasample_filename = data/SA_test_data/strat_tables/strat_$collarID$.csv
dist_table_filename = data/SA_test_data/dh_asud_strat2.csv

[DataHeaders]
drillsample_header = DEPTH_FROM_M, DEPTH_TO_M, MAJOR_LITHOLOGY,
stratasample_header = DEPTH_FROM_M, DEPTH_TO_M, STRAT_UNIT_NAME,
strata_data_header = strat, summary, distance, description

[SolverParameters]
add_topology_constraints = True
max_num_strata_jumps = 0
max_num_returns_per_unit = 0
max_num_unit_contacts_inside_litho = 0
single_top_unit = True

[Correlation]
correlation_power = 1.0

[DataPreprocessing]
number_nearest_units = 10
min_drillhole_litho_score = 80
group_drillhole_lithos = False
cover_ratio_threshold = 0.65

[CollarIDs]
collarIDs = 265003,265010,265018,265030,265043,265051

```

https://doi.org/10.5194/egusphere-2025-1294
Preprint. Discussion started: 26 May 2025

Alvarado-Neves, F., Ailleres, L., Grose, L., Cruden, A. R., & Armit, R. (2024). Three-dimensional geological modelling of igneous intrusions in LoopStructural v1.5.10. *Geoscientific Model Development*, *17*(5), 1975–1993. https://doi.org/10.5194/gmd-17-1975-2024

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

Caumon, G., Collon-Drouaillet, P., Le Carlier de Veslud, C., Viseur, S., & Sausse, J. (2009). Surface-Based 3D Modeling of Geological Structures. *Mathematical Geosciences*, *41*(8), 927–945. https://doi.org/10.1007/s11004-009-9244-2

D'Affonseca, F. M., Finkel, M., & Cirpka, O. A. (2020). Combining implicit geological modeling, field surveys, and hydrogeological modeling to describe groundwater flow in a karst aquifer. *Hydrogeology Journal*, *28*(8), 2779–2802. https://doi.org/10.1007/s10040-020-02220-z

Giraud, J., Pakyuz-Charrier, E., Jessell, M., Lindsay, M., Martin, R., & Ogarko, V. (2017). Uncertainty reduction through geologically conditioned petrophysical constraints in joint inversion. *GEOPHYSICS*, *82*(6), ID19–ID34. https://doi.org/10.1190/geo2016-0615.1

Guo, J., Wang, Z., Li, C., Li, F., Jessell, M. W., Wu, L., & Wang, J. (2022). Multiple-Point Geostatistics-Based Three-Dimensional Automatic Geological Modeling and Uncertainty Analysis for Borehole Data. *Natural Resources Research*, *31*(5), 2347–2367. https://doi.org/10.1007/s11053-022-10071-6

Guo, J., Xu, X., Wang, L., Wang, X., Wu, L., Jessell, M., Ogarko, V., Liu, Z., & Zheng, Y. (2024). GeoPDNN 1.0: a semi-supervised deep learning neural network using pseudo-labels for three-dimensional shallow strata modelling and uncertainty analysis in urban areas from borehole data. *Geoscientific Model Development*, *17*(3), 957–973. https://doi.org/10.5194/gmd-17-957-2024

Hagberg, A. A., Schult, D. A., & Swart, P. J. (2008). *Exploring Network Structure, Dynamics, and Function using NetworkX*. 11–15. https://doi.org/10.25080/TCWV9851

Hartmann, J., & Moosdorf, N. (2012). The new global lithological map database GLiM: A representation of rock properties at the Earth surface. *Geochemistry, Geophysics, Geosystems*, *13*(12). https://doi.org/10.1029/2012GC004370

Himsolt, M. (1997). *GML: a portable graph file format*.

Jessell, M. (2001). Three-dimensional geological modelling of potential-field data. *Computers & Geosciences*, *27*(4), 455–465. https://doi.org/10.1016/S0098-3004(00)00142-4

Jessell, M., Aillères, L., Kemp, E. de, Lindsay, M., Wellmann, F., Hillier, M., Laurent, G., Carmichael, T., & Martin, R. (2014). Next Generation Three-Dimensional Geologic Modeling and Inversion. In *Building Exploration Capability for the 21st Century*. Society of Economic Geologists. https://doi.org/10.5382/SP.18.13

Jessell, M., Ogarko, V., de Rose, Y., Lindsay, M., Joshi, R., Piechocka, A., Grose, L., de la Varga, M., Ailleres, L., & Pirot, G. (2021). Automated geological map deconstruction for 3D model construction using <i>map2loop</i> 1.0 and <i>map2model</i> 1.0. *Geoscientific Model Development*, *14*(8), 5063–5092. https://doi.org/10.5194/gmd-14-5063-2021

https://doi.org/10.5194/egusphere-2025-1294
Preprint. Discussion started: 26 May 2025

Jessell, M. W., Ailleres, L., & de Kemp, E. A. (2010). Towards an integrated inversion of geoscientific data: What price of geology? *Tectonophysics*, *490*(3–4), 294–306. https://doi.org/10.1016/j.tecto.2010.05.020

Joshi, R., Madaiah, K., Jessell, M., Lindsay, M., & Pirot, G. (2021). <i>dh2loop</i> 1.0: an open-source Python library for automated processing and classification of geological logs. *Geoscientific Model Development*, *14*(11), 6711–6740. https://doi.org/10.5194/gmd-14-6711-2021

Land, A. H., & Doig, A. G. (1960). An Automatic Method of Solving Discrete Programming Problems. *Econometrica*, *28*(3), 497. https://doi.org/10.2307/1910129

Lindsay, M. D., Jessell, M. W., Ailleres, L., Perrouty, S., de Kemp, E., & Betts, P. G. (2013). Geodiversity: Exploration of 3D geological model space. *Tectonophysics*, *594*, 27–37. https://doi.org/10.1016/j.tecto.2013.03.013

Maldonado, A., & Mercer, K. (2018). *Comparison of the Laboratory and Barton-Bandis Derived Shear Strength of Bedding Partings in Fresh Shales of the Pilbara, Western Australia*. *All Days*, ISRM-ARMS10-2018-204.

Mallet, J.-L. L. (2002). *Geomodeling*. Oxford University Press, Inc.

Martin, R., Ogarko, V., Giraud, J., Plazolles, B., Angrand, P., Rousse, S., & Macouin, M. (2024). Gravity data inversion of the Pyrenees range using Taguchi sensitivity analysis and ADMM bound constraints based on seismic data. *Geophysical Journal International*, *240*(1), 829–858. https://doi.org/10.1093/gji/ggae410

Ogarko, V., Giraud, J., Martin, R., & Jessell, M. (2021). Disjoint interval bound constraints using the alternating direction method of multipliers for geologically constrained inversion: Application to gravity data. *GEOPHYSICS*, *86*(2), G1–G11. https://doi.org/10.1190/geo2019-0633.1

Ogarko, V., & Jessell, M. (2025). litho2strat 1.0 source code, https://doi.org/10.5281/zenodo.15064469

Pakyuz-Charrier, E., Giraud, J., Ogarko, V., Lindsay, M., & Jessell, M. (2018). Drillhole uncertainty propagation for three-dimensional geological modeling using Monte Carlo. *Tectonophysics*, *747–748*, 16–39. https://doi.org/10.1016/j.tecto.2018.09.005

Schetselaar, E. M., & Lemieux, D. (2012). A drill hole query algorithm for extracting lithostratigraphic contacts in support of 3D geologic modelling in crystalline basement. *Computers & Geosciences*, *44*, 146–155. https://doi.org/10.1016/j.cageo.2011.10.015

Tarantola, A. (2005). *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9780898717921

Vollgger, S. A., Cruden, A. R., Ailleres, L., & Cowan, E. J. (2015). Regional dome evolution and its control on ore-grade distribution: Insights from 3D implicit modelling of the Navachab gold deposit, Namibia. *Ore Geology Reviews*, *69*, 268–284. https://doi.org/10.1016/j.oregeorev.2015.02.020

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