

Peer Review Report

Review Report on Geospatial data and deep learning expose ESG risks to critical raw materials supply: The case of lithium

Original Research, Earth Sci. Syst. Soc.

Reviewer: Matthew Cracknell

Submitted on: 05 Apr 2024

Article DOI: 10.3389/esss.2024.10109

EVALUATION

Q 1 Please summarize the main findings of the study.

See attached report

Q 2 Please highlight the limitations and strengths.

See attached report

Q 3 Please comment on the methods, results and data interpretation. If there are any objective errors, or if the conclusions are not supported, you should detail your concerns.

See attached report

Q 4 Check List

Is the English language of sufficient quality?

Yes.

Is the quality of the figures and tables satisfactory?

Yes.

Does the reference list cover the relevant literature adequately and in an unbiased manner?

Yes.

Are the statistical methods valid and correctly applied? (e.g. sample size, choice of test)

Yes.

If relevant, are the methods sufficiently documented to allow replication studies?

Yes.

Are the data underlying the study available in either the article, supplement, or deposited in a repository? (Sequence/expression data, protein/molecule characterizations, annotations, and taxonomy data are required to be deposited in public repositories prior to publication)

Yes.

Does the study adhere to ethical standards including ethics committee approval and consent procedure?

Not Applicable.

If relevant, have standard biosecurity and institutional safety procedures been adhered to?

Not Applicable.

Q 5 Please provide your detailed review report to the editor and authors (including any comments on the Q4 Check List):

Manuscript ID 10109 – peer review

Geospatial data and deep learning expose ESG risks to critical raw materials supply: The case of lithium
5 April 2024

The manuscript “Geospatial data and deep learning expose ESG risks to critical raw materials supply: The case of lithium” explores the impact of environmental, social, and governance (ESG) factors on the supply chains of critical raw materials (CRM), using public ESG data and a global dataset of news events to predict natural resources conflicts. Three models were developed, with the deep learning model demonstrating the highest accuracy, but also revealing biases influenced by human settlements and their proxies. The study suggests that mapping spatially situated ESG risks for different geological sources of CRM can contribute to sustainable land-use planning, investing, and policy-making.

The manuscript is well written and structured and presents a valuable and novel contribution to the geoscience community by incorporating ESG through modelling risk of natural resources conflict. The authors have used robust input data and methods to generate three models (knowledge- and data-driven). In my view this research should be published after addressing a few minor edits and comments to improve clarity (provided below and in the general comments section).

I am a little confused with the concept of ESG rating. I realise after reading this a couple of times that the ESG rating specifically refers to the output of the Fuzzy logic model. This ESG rating is then not entirely comparable to the Naïve Bayes and Deep Learning model outputs (which is the probability of natural resource conflict). This distinction needs to be stated more clearly to the reader.

The authors need to clearly state that the ESG rating from the Fuzzy logic model is linked to Li occurrence (and not the NB or DL model outputs) and why. I suggest stating this in the abstract and in the results section on Li mineral occurrence at a bare minimum.

The discussion around the distinct benefits of the knowledge-driven (less influenced by spatial bias of input data) and data-driven, specifically DL, model (characterising non-linear relationships between inputs for feature importance interpretation) will benefit from short unifying statements in the abstract and conclusions.

General comments

Abstract

Lines 33–34: It is unclear that the ESG risk model and the lithium mineral systems model are essentially sequential steps (ESG risk model first then linked to Li occurrence). This needs to be specified here and in the introduction and method and conclusions.

Introduction

Line 111: Please clarify that lithium is the case study example in this paper. At the moment the wording of the sentence starting with “For instance ...” is a little vague to the reader. I suggest something like “in this study Li ...”

Line 112: Do you need to describe what a mineral system is?

Line 131: Please specify which models are the “more advanced machine learning models”, i.e. models 2 and 3 (if this is the case ... although the wording in Line 126 implies that the Naïve Bayes mode is simplistic compared to the Deep Learning model).

Line 132: This sentence is the first mention that the ESG model and the Li model are separate, e.g. the ESG model evaluated separately and is subsequently used to inform the Li occurrence model. This needs to be stated clearly in the abstract. See comment above.

Line 133–134: suggest rewording last sentence to “Despite a focus on Li, our ESG models are applicable to other CMR ...”

Data

Line 162: 39 Mt of Li₂O? The zeros seem cumbersome.

Line 167: “Igneous mineral systems ...” to be consistent with following paragraphs. I would prefer to highlight each of the minerals systems descriptions using sub headers (could be cumbersome) or highlighting via italics ... but this is my personal preference.

Line 246 and other places where “discussed below”: please indicate the specific section of the discussion that contains the related comments.

Lines 323–324 and 329: Why the general decrease in filtered POLECAT news stories with time? Is this related to the removal of duplicates?

Line 326: missing parenthesis

Line 355: a little off topic here ... there is contention over importance of petroleum resources as a motivator in the “oil wars” (e.g. Meierding, E. 2020 *The Oil Wars Myth: Petroleum and the Causes of International Conflict*. Cornell University Press.). Perhaps this sentence needs to be more carefully framed around natural resources in general?

Methods

Line 402: Do you incorporate fold 6 into the knowledge-driven fuzzy logic model? Please specify

Line 421 onwards: Naïve Bayes requires approximately gaussian distributed predictor variables. Despite the relative success of this method was normalisation applied to the inputs?

Results

Lines 509 and others: Here you state that higher values from the fuzzy logic model are interpreted as an increase in “ESG rating”, which I interpret to mean that there are lower natural resources conflict risk. For the other models (naïve Bayes and DL) the output metric seems to be different, described as the probability of natural resource conflict, which is interpreted as high values represents high risk. If I have interpreted this correctly it is confusing. I may have missed the section that clarifies the reasoning and need for two (inverse) metrics to quantify natural resource conflict.

Line 511: remove extra parenthesis.

Lines 526+: Please clarify that only the fuzzy logic ESG conflict model is linked with the Li occurrences data (to generate Li ESG ratings) and that the other ESG conflict models have not been. You may need a lead in sentence that justifies the reasoning for only choosing the fuzzy logic model for this task, e.g. spatial bias in the data-driven models (there is a hint at this in Lines 685–687).

Discussion

Line 595: delete extra parenthesis

Lines 659–652: I understand the reasoning behind the use of the Fuzzy logic model for the Li case study. However, here you are sprucing the benefits of the DL model in terms of reporting feature importance. I feel that the distinct and unique benefits and limitations of the knowledge- and data-driven models have not been clearly stated (I am still having trouble reconciling them after reading this a couple of times). Not sure of specifically where this needs to occur, perhaps if clarified in the abstract and in the introduction?

Line 674: quotation marks around friend-shoring?

Line 677 and others: capitalise “arctic”.

Lines 770–771: fair point. This would be very interesting to build a spatio-temporal models of ESG risk and see how it evolves over time for a given jurisdiction. This would db e very helpful for forecasting CRM opportunities.

Line 786–787: Any suggestions on for gathering/acquiring these more granular data? Could feature importance rankings offer guidance on which data to focus on to improve the models most efficiently?

Line 803: this statement around the robustness of knowledge-driven models needs to be stated in the abstract.

Figures

Figure 1. Deel Learning should be Deep Learning

Figure 7. Inset box locations on global map might be helpful?

Figure 9. Need to state in the caption that the ESG ranking is specifically generated from the fuzzy logic model.

Figure 10. Please specify what the orange areas represent (protected areas?) either in the figure or the caption.

References

Line 1248: I note that Owen et al. 2022a, is just a dataset. Is there supporting documentation?

QUALITY ASSESSMENT

Q 6 Originality



Q 7 Rigor



Q 8 Significance to the field



Q 9 Interest to a general audience



Q 10 Quality of the writing



Q 11 Overall quality of the study

