essd-2022-418

Responses to Reviewer 1

This manuscript proposed to create a digital soil map of Australia based on a series of environmental covariates acquired at distinct spatial resolutions. Modeling was achieved with a machine learning algorithm trained on a large available soil geochemistry dataset and evaluated in another independent soil dataset. While the premise of the work seems very interesting and the results provided can be a major contribution to mineral exploration in Australia, I have some concerns regarding the representability of the input data after resampling as well of the external validation dataset. The correlation between input and predicted variables is also not very convincing. Some key literature works could help to give another dimension to the work, especially in the discussion. I would also like to see some general comments on the possibility of conducting similar approaches in other regions of the globe. I will try to explain each of these issues in more detail, hoping to help the authors to improve their work.

Despite the low prediction accuracy, this paper demonstrates a step forward in the development of machine learning in generating predictive geochemical maps, not limited to Li. Similar approach can be implemented globally provided the availability of national database of the geochemical data along with relevant environmental covariates.

Specific comments

Main comments

1. It should be clear from the beginning of the abstract which other types of input data (besides geochemical data) were used for modeling.

We did mention that we utilised both data from geochemical survey and environmental covariates to predict and map the spatial distribution of Lithium. We further clarify in the text that the environmental covariates that affect soil formation are utilised.

Add input data

2. One of my main concerns is related to model validation since the authors claim the success of the method proposed based on the external validation dataset. Is NAGS representative of all of Australia? In my view, soil characteristics and the Li content of the soils will vary throughout the country, and model performance will vary accordingly. Moreover, the NGSA and NAGS correspond to different sampling campaigns, with different collection dates and sampling densities. Despite the leveling employed by the authors, how can you confirm that the NAGS is suitable for the validation of the model trained in another dataset? Are the results obtained with the NAGS comparable to the out-of-the-bag validation using the NGSA data?

We agree that the model performance would be different for different sites and understand that NAGS may not be representative of all Australia. It is the best available data that we can obtain to evaluate the model performance. We believe it is more suitable for evaluation, instead of simply dividing the NGSA dataset as calibration and validation. NAGS is suitable for the validation of the model trained in another dataset because the distribution of the data is similar based on the Kolmogorov-Smirnov test. The validation results from NAGS ($R^2 = 0.32$) are higher than those for NGSA TOS ($R^2 = 0.20$).

3. Resampling of the data: the authors have resampled data acquired at 30 m or 90 m spatial resolution to a final resolution of 3 km. How can you ensure important information is not being lost with the resampling?

A bilinear resampling method is implemented. As this is a continental evaluation, we are aiming to provide a continental trend rather than local trend.

4. Remote sensing has limited penetration depth (0-10cm). How can you correlate the remote sensing variables with the BOS dataset?

Rao, K.S., Chandra, G. & Narasimha Rao, P.V. Study on penetration depth and its dependence on frequency, soil moisture, texture and temperature in the context of microwave remote sensing. J Indian Soc Remote Sens 16, 7–19 (1988). https://doi.org/10.1007/BF03014300

We agree that the remote sensing (Landsat) has limited penetration depth. In this study, we trained a prediction model for each depth, and developed the correlation of remote sensing data and the elemental concentration from each depth separately. We did not solely use Landsat data, we also used climate, subsoil clay content and topographical data. In addition, the BOS data is highly correlated with TOS ($r = 0.75$, $p < 0.0001$).) and thus we can still infer BOS from climate, soil, topography and Landsat information.

1. Figure 3: all variables show correlations below 0.3, which is considered by many authors as a negligible correlation. Taking this into account, how can you create a reliable model and consequently digital soil map?

We did mention in the text that the correlation of an individual variable to Lithium concentration is relatively low. We did not claim that we developed reliable model. This manuscript demonstrates a case study on utilisation of machine learning to develop predictive geochemical maps.

2. Which is the advantage of this method compared with traditional interpolation approaches (IDW, kriging)? I am aware that a comparison with other methods is beyond the scope of this manuscript, but the authors could comment on the advantages/disadvantages compared to previous works (if available), considering the NGSA dataset.

Both IDW (Inverse Distance Weight) and kriging are interpolation method that utilise known values at certain location to predict values at unknown points without the use of covariates. The DSM method predicts the unknown values from the trained chemometrics model that relate the environmental covariates with the known values at each depth using machine learning models.

3. Similarly, which is the advantage of the produced maps (Figure 6) when compared with the maps of Figure 1? Are any regions highlighted by the proposed method that were not highlighted in Figure 1? This information would be important for the readers to assess with the method you proposed is of interest.

Figure 1 represents the Li concentration at a particular location, while Figure 6 represents the predictive Li concentration across the map. We point out locations that have anomalous Li concentration that can be investigated. Their shape and extent are defined using the higher spatial resolution model than the original sparse point-data.

Minor comments

Introduction

- Lines 56-60: the authors explain that Li extraction from brines is in the form of Li-chloride. However, it would also be important to clarify that Li-carbonate is not directly extracted from Li-minerals, but the Li metal instead.

Thank you for the comment. We added this into the revised version of the manuscript.

- Lines 61-76: a brief description of the behavior of Li in soils is presented and previous works on soil geochemistry are presented. Other works related to this topic are also worth mentioning:

Luecke, W. (1984). Soil Geochemistry above a Lithium Pegmatite Dyke at Aclare, Southeast Ireland. Irish Journal of Earth Sciences, 6(2), 205–211. http://www.jstor.org/stable/30002472

Steiner, B. (2018). Using Tellus stream sediment geochemistry to fingerprint regional geology and mineralisation systems in Southeast Ireland. Irish Journal of Earth Sciences, 36, 45-61. doi: 10.3318/ijes.2018.36.45.

We have included the additional references as suggested.

- Lines 82-89: a summary of mineral prospectivity mapping is made, but other recent prospectivity mapping studies are missing:

Parsa, M. (2021). A data augmentation approach to XGboost-based mineral potential mapping: An example of carbonate-hosted ZnPb mineral systems of Western Iran. Journal of geochemical exploration, 228, 106811. doi: https://doi.org/10.1016/j.gexplo.2021.106811.

von der Heyden, B. P., Todd, C., Mayne, M. J., & Doggart, S. (2023). Zipf rank analysis highlights the exploration potential for Lithium-Caesium-Tantalum -type pegmatites in the Northern Cape, South Africa. Journal of African Earth Sciences, 197, 104769. doi: https://doi.org/10.1016/j.jafrearsci.2022.104769.

We added more recent references as suggested.

- Lines 92-96: a short literature review on the use of remote sensing for Li pegmatite identification is presented. Some of these works could be replaced by more recent studies:

Cardoso-Fernandes, J., Teodoro, A. C., Lima, A., & Roda-Robles, E. (2020). Semi-Automatization of Support Vector Machines to Map Lithium (Li) Bearing Pegmatites. Remote Sensing, 12(14), 2319. doi: 10.3390/rs12142319.

Morsli, Y., Zerhouni, Y., Maimouni, S., Alikouss, S., Kadir, H., & Baroudi, Z. (2021). Pegmatite mapping using spectroradiometry and ASTER data (Zenaga, Central Anti-Atlas, Morocco). Journal of African Earth Sciences, 177, 104153. doi: https://doi.org/10.1016/j.jafrearsci.2021.104153.

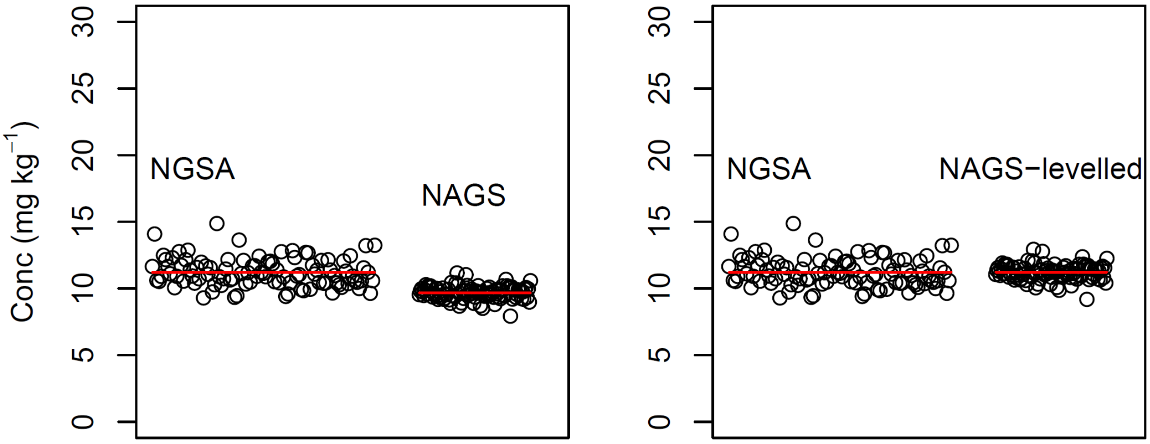
Booysen, R., Lorenz, S., Thiele, S. T., Fuchsloch, W. C., Marais, T., Nex, P. A. M., & Gloaguen, R. (2022). Accurate hyperspectral imaging of mineralised outcrops: An example from lithium-bearing pegmatites at Uis, Namibia. Remote Sensing of Environment, 269, 112790. doi: https://doi.org/10.1016/j.rse.2021.112790.

We added more recent references as suggested.

Materials and methods

- Lines 143-145: “levelling method were utilized using the standards Certified Reference Materials (Main and Champion, 2022). In short, a correction factor based on the CRM measurements from the two datasets is calculated and applied as multiplier to relevel the data”. Can we see a figure showing this leveling process?

The following figure is for illustrative purpose only as no actual data was made available. Only the average of CRM-TILL 1 measurement was provided for each dataset.



The figure on the left showed the spread of TILL-1CRM from both NGSA and NAGS dataset. It can be clearly seen that the concentration of TILL-1 CRM on NAGS dataset is slightly lower than those from NGSA. The correction factor is developed by calculating the ratio of mean concentration from NGSA dataset against the NAGS dataset. The NAGS is then levelled using the correction factor for better representation. This is very important because NGSA dataset was utilised for the prediction on the NAGS dataset.

- Table 1: Spatial resolution of Landsat data is 30 m and not 25 m.

We are not utilising raw Landsat data. The Landsat data we used was produced from another study (Wilford and Roberts, 2019) that pre-processed the Landsat data to expose Australia on its barest land.

Results and discussion

- Line 239: "concentrations ranging from 0.1 – 67.4 and 0.1 – 56 mg kg-1, for TOS and BOS respectively". These values do not seem to match Figure 2. Please revise.

The bin width was too large in the original figure. We have corrected this.

- Line 242: "mean Li concentration". Which is the mean for the TOS and BOS datasets? Right now, you are comparing the median and the mean.

We have corrected this.

- Lines 261-262: "Despite other studies (Robinson et al., 2018; Kashin, 2019) reporting strong correlations between Li and Mg, and other elements elsewhere". Extensive work on Li correlations with other elements for both stream sediment samples and a large pegmatite dataset for the Iberian Peninsula. Please consider comparing your results with other works.

Cardoso-Fernandes, J., Lima, J., Lima, A., Roda-Robles, E., Köhler, M., Schaefer, S., Barth, A., Knobloch, A., Gonçalves, M. A., Gonçalves, F., & Teodoro, A. C. (2022).

Stream sediment analysis for Lithium (Li) exploration in the Douro region (Portugal): A comparative study of the spatial interpolation and catchment basin approaches. Journal of geochemical exploration, 236, 106978. doi: https://doi.org/10.1016/j.gexplo.2022.106978.

We have provided the comparison of our results to different studies.

- Section 3.1.1: Please explain which threshold was used to consider a strong/moderate correlation.

We have added the definition of threshold for correlation analysis.

- Lines 267-271: Please consider key works on the alteration of Li minerals to clays:

London, D., & Burt, D. M. (1982). Chemical models for lithium aluminosilicate stabilities in pegmatites and granites. American Mineralogist, 67(5-6), 494-509.

Quensel, P. (1937). Minerals of the Varuträsk Pegmatite. Geologiska Föreningen i Stockholm Förhandlingar, 59(2), 150-156. doi: 10.1080/11035893709444939.

Quensel, P. (1938). Minerals of the Varuträsk Pegmatite. Geologiska Föreningen i Stockholm Förhandlingar, 60(2), 201-215. doi: 10.1080/11035893809444995.

We have added additional references as suggested.

- Line 283: “Landsat bands 3, 5 and 6 had stronger negative correlations (r = -0.14 to -0.16)”. Please notice that -0.15 and -0.17 represent stronger negative correlations than -0.14 and -0.16. Moreover, the graph scale is not the same for TOS and BOS in Figure 3. That is why the bars seem bigger for the BOS data when the values are smaller in module.

We have rescaled the x axis to be the same, and reworded the sentence.

- Lines 306-307: there are some studies on the spectral behavior of Li minerals and in some cases cross-analysis with the Li content:

Cardoso-Fernandes, J., Silva, J., Perrotta, M. M., Lima, A., Teodoro, A. C., Ribeiro, M. A., Dias, F., Barrès, O., Cauzid, J., & Roda-Robles, E. (2021). Interpretation of the Reflectance Spectra of Lithium (Li) Minerals and Pegmatites: A Case Study for Mineralogical and Lithological Identification in the Fregeneda–Almendra Area. Remote Sensing, 13(18), 3688. doi: 10.3390/rs13183688.

I am missing the point here. The study refers to the use of reflectance spectroscopy to identify types of mineral deposits; this is more towards proximal sensing instead of mapping despite the authors claiming similar concept can be applied to the reflectance spectroscopy data obtained from satellite data (which most likely will be noisier).

- Lines 334-335: “the model separates out prediction values based on its spectral response of vegetation”. I do not understand. Didn't you use the bare soil dataset where the vegetation influence was removed? Please comment.

The bare soil dataset does not remove data for annual vegetation. It mainly removed seasonal vegetation influence. We double checked the representation of model, and noted a mistake was made, and have corrected the figure.

- Line 343: “Landsat bands 2 and 6, and temperature range also affect model conditions”. Again, remote sensing data has low penetration depth, therefore the correlation with the BOS dataset should be low. How do you explain these results?

Not applicable. We double checked the representation of model, and noted a mistake was made, and have corrected the figure. Please see the reply above on the correlation between TOS and BOS.

- Lines 356-357: “the model developed here to have a higher concentration of soil Li, especially for the BOS model”. However, this is the model with a higher standard deviation. Please comment.

We simply claim that the predictive map yielded results that make sense. There is higher Li concentration on areas where the existing Li mines were located.

- Figure 7 is just a zoom of Figure 6, not bringing new information. I would prefer a comparison between the predicted contents in the validation area with both NGSA and NAGS measured values.

Figure 7 is indeed a zoom of Figure 6. It might be of interest to observe the variation within an area.

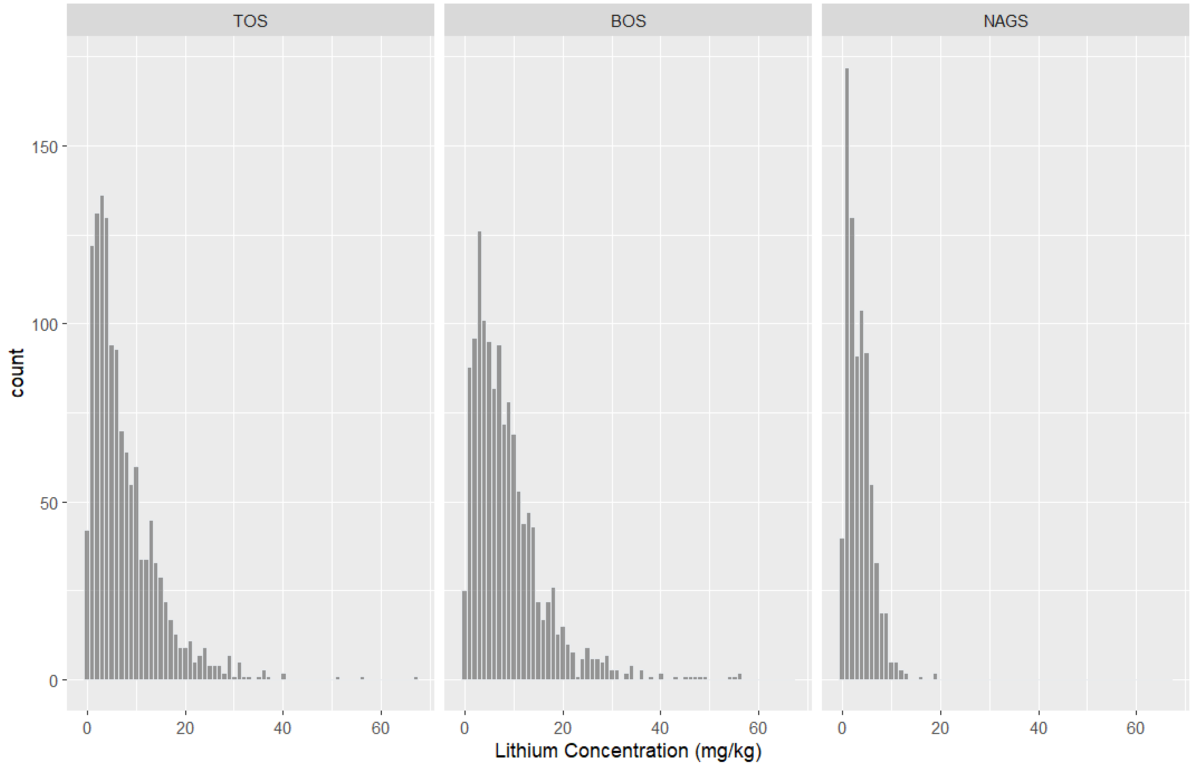
Technical corrections

- Figure 1: Can you improve the quality of Figure 1? Is it possible to display the Li concentration with a ramp color to aid visualization? Can you separate the two maps into subfigures A (TOS) and B (BOS)?

We have improved the quality of the figure for review. It is possible to do Lithium concentration with ramp color. However, we believed it would make the map too noisy. Hence, we kept the previous way of plotting. We would like to keep both figures as a set for cohesion purpose.

- Figure 2: in the histogram of the left we don't see values > 40 mg/kg. Also, can we see the histogram for the NAGS dataset?

We have included the histogram for all dataset



- Figure 3: the graph bars go further than the X-axis. Please correct this issue.

We have updated the plot.

- Line 301: “Higher accuracy was observed in TOS”. Higher, but still low. Please consider revising the sentence.

We have corrected the sentence.

- Table 2: the metric values presented for the external validation do not match the values mentioned in the text. Please revise.

We have revised the values.

- Figure 8: there are no units on the Y-axis.

We have added units on the Y-axis.

- Please revise the use of acronyms throughout the manuscript.

We have revised the use of acronyms throughout the manuscript.

Please see the **attached pdf file** (edited version of the original file) with some minor corrections/suggestions and yellow highlights that need to be addressed carefully.

essd-2022-418

Responses to Reviewer 2

This manuscript presents results from a study designed to develop predictive maps of the Li concentration in soil or Australia. The calibration data set used was the Li data generated during the National Geochemical Survey of Australia. In addition, several environmental covariates were used such as annual precipitation, annual evaporation, airborne radiometric data, etc. The predictive Li maps were then compared to a validation data set from the northern Australia Geochemical Survey. The paper is relatively well-written and well organized. Unfortunately, the results of the study were disappointing in that the correlation between observed Li in the validation data ser and the predicted Li values from the model was relatively poor. The authors have recognized several limitations from their study, but have neglected to discuss what I think is an important issue—the nature of the Li data from the NGSA.

We thank the reviewer for the constructive comments. We have clarified within the revised manuscript that the Lithium concentration mentioned in the text are aqua regia extractable Li.

The authors should emphasize that they are using aqua-regia extractable Li data from the NGSA in this study. Aqua-regia digestion only extracts a portion of the total Li found in soil. I am not sure what the fraction of the total that may be, but it depends on the extraction parameters (e.g., temperature of extraction, length of time the soil material is left in contact with the aqua regia) as well as soil mineralogy. Lithium in clay minerals may be extracted, but I am not sure that Li in spodumene will be released by aqua regia extraction. The national-scale soil geochemical survey of the conterminous United States (Smith et al., 2019, complete reference given below) used a 4-acid extraction that is a much more vigorous extraction than aqua regia and should give a good estimate of the total Li content in soil For this US study, three samples were collected at each site (4,857 sites): soil from a depth of 0-5 cm. a composite of the soil A horizon, and a sample of the top 20 cm of the soil C horizon. The results shows a range of <1-300 mg/kg (median 20 mg/kg) for the 0-5 cm sample; a range of <1-315 mg/kg (median 20 mg/kg) for the soil A horizon; and a range of <1-280 mg/kg (median 24 mg/kg) for the soil C horizon. These concentrations are considerably higher than the aqua regia extraction data for Australia. So one might ask if the results of the current study would be different if total Li data were used instead of the aqua-regia-extractable data? Another question is whether a weaker extraction that only released "plant available" Li might be more likely to give better results.

We thank the reviewer for the constructive comments and pointed out the nature of the NGSA sample and the possibility of better results if Total Li was measured. We have included this in comparison with the NGSA data in our revised paper.

Despite the somewhat disappointing results of this study, I feel the paper should be published to demonstrate a step forward in the development of machine learning in generating predictive geochemical maps. The authors should also perhaps do a better job of recognizing the importance of the National Geochemical Survey of Australia whose data gave them the opportunity to conduct the current study. Perhaps a recommendation might be to recognize the need to conduct higher density national- and international-scale geochemical surveys and to add additional parameters to these studies (e.g., quantitative mineralogy) that would aid in future studies such as these authors have conducted.

Reference for Smith et al. (2019):

Smith, D.B., Solano, F., Woodruff, L.G., Cannon, W.F. and Ellefsen, K.J. (2019). Geochemical and Mineralogical Maps, with Interpretation, for Soils of the Conterminous United States. *United States Geological Survey Scientific Investigations Report, 2017-5118,* https://pubs.usgs.gov/sir/2017/5118/index.html.

We thank the reviewer for the comments and added this in our discussion.

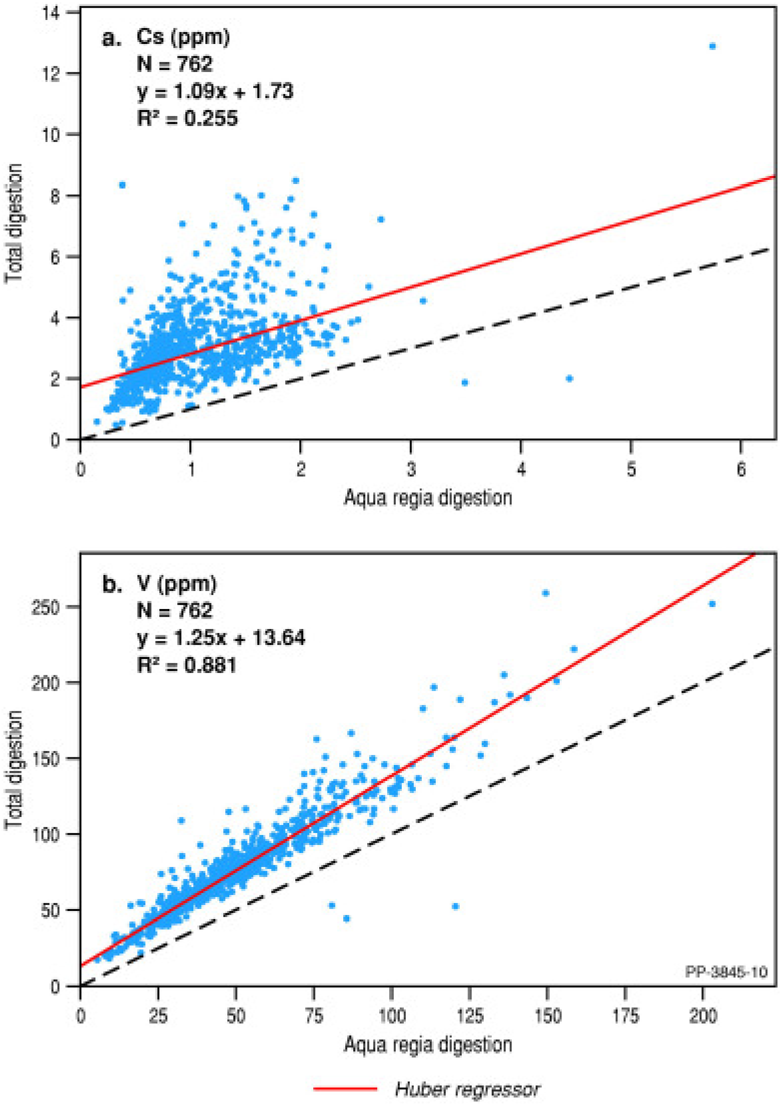
Prior to ICP-MS analysis, the samples underwent the following digestion process:

A 0.50 ± 0.02 g aliquot of the sample was digested with an aqua regia solution (1.8 mL concentrated hydrochloric acid added first followed by 0.6 mL concentrated nitric acid) at 90 ± 3 °C for 2 hours to leach the acid-soluble components (the leach only partially dissolves silicates and oxides). Once the sample had cooled to room temperature, 17.5 mL of diluent was added and the bottle was capped and inverted 10 times to completely homogenise the content. The sample was then allowed to settle over a 6 hour period or longer. The sample was exposed to light for as little time as possible to prevent reduction of silver. The sample was diluted a further 50 times with 18 MΩ/cm water.

We have added this information within the manuscript. We also mentioned within the text (Li measurement section) the any reference to Li refers to aqua regia extractable Li. The reason total Li was not measured in the NGSA (or NAGS) samples is simply that all total analyses were done on XRF fused beads made using Li-borate flux.

We are not sure how much the aqua-regia extractable portion accounts for in comparison to the total Li because the geochemical survey data for Lithium was only conducted on the aqua regia extractable portion. Depending on the element, the concentration of aqua regia digestion could vary significantly in comparison to those from total digestion. (see Figure below from Main and Champion (2022)).

Main, P. T. and Champion, D. C.: Levelling of multi-generational and spatially isolated geochemical surveys, Journal of Geochemical Exploration, 240, doi::1016/j.gexplo.2022.107028, 2022.



The results could potentially be improved if the total Li was measured (this would require a completely new analytical process). Most of the observations collected had relatively low concentration; having more representative samples at higher concentration might improve the prediction accuracy.

Thank you for the suggestions. We have added those in the conclusion section.

I have made a few specific suggestions and editorial comments below:

1. Lines 16-17: The authors begin this sentence by stating that “soil samples were collected.” Then later in the sentence, they refer to “catchment outlet sediments.” Are they saying that all the catchment outlet sediments sampled during the National Geochemical Survey of Australia can also be considered

to be “soils?” This is explained later in the text (lines 121-124); however, it probably deserves a sentence in the abstract to clarify this issue. An alternative would be to avoid the use of “catchment outlet sediments” in the abstract.

We have corrected it into catchment outlet sediments samples.

2. Lines 24-25. Note this sentence: “The map shows high Li concentration around existing mines and other potentially anomalous Li areas.” It seems a bit strange to this reviewer to say there are high Li concentrations around potentially anomalous Li areas. If there are high Li values, then the area is, by definition, anomalous. Perhaps the sentence is constructed in this manner because the map to which the authors refer is predictive and it would require collecting physical samples from the “potentially anomalous” areas to confirm if they were actually anomalous.

   We have corrected the sentence into:

   The predicted maps show high Li concentration around existing mines and other potentially anomalous Li areas that have yet to be verified.

3. Line 43: Change “was the second” to “is second”; change “the first” to “first”

   We have corrected the words.

4. Line 44: Change “economic resource” to “economic resources.” Change “According to recent survey” to “According to a recent survey”

   We have corrected the words.

5. Line 47: Change “Li is hosted mainly spodumene” to “Li is hosted mainly in spodumene” Change “while” to “whereas”

   We have corrected the words.

6. Line 142: Note this sentence: “Furthermore, these samples were collected at different times and /or laboratories.” Do you mean that multiple laboratories were used to analyze the samples? It is unclear what exactly is meant here.

   Samples from NAGS dataset were collected and analysed at different laboratories to those samples from NGSA dataset. We have clarified the sentence.

7. Lines 142-145. In my opinion, there should be a more specific discussion about the leveling of these two data sets. What Certified Reference Material was used? Was it Till-1 as mentioned previously for the NGSA data? How

about a simple plot of NGSA Li concentration versus NAGS Li concentrations to give the reader a better idea of data comparability. I just do not think there is sufficient information given here.

The TILL-1 CRM is used. We clarified the process taken to level the dataset in the manuscript. We cannot compare the NGSA and NAGS concentration directly as the measurements were taken at different locations.

8. Lines 148-149. Change ", that contributes" to "that contribute"

   We have corrected the word to "that contribute".

9. Line 178. Change "map" to "mapping" or just omit the word "map"

   We have removed the word "map".

10. Line 179-180. I think there is a word missing between "measurements" and "soil" in line 179. Perhaps it should be "measurements on soil?"

    We have corrected the sentence.

11. Line 238. I would suggest using "aqua-regia-extractable Li concentrations" instead of just saying "Li concentrations" here and anywhere else in the text. This lets the reader know that you are using data derived from a partial extraction (aqua regia) and not the total Li content of the soil. Another option would be to have a sentence early in the text to state that for the remainder of the paper, any reference to Li concentrations is understood to mean aqua-regia-extractable Li unless otherwise noted.

    We have already mentioned that the Li concentration is derived from the aqua-regia digestion method (Line 137-138). We re-emphasised it again later on within the same paragraph.

12. Line 242. Here, again, I would suggest using "aqua-regia extractable Li concentrations" when referring to the Negrel et al. (2019) publication. The European study also used an aqua regia extraction in the determination of Li concentrations, so their data should be comparable to the Australian data.

    We have updated the sentence

13. Line 243. I do not know what data Schrauzer (2002) used for obtaining the estimated range of 7-200 mg/kg for a world background Li concentration. If you have that information, I would suggest including it in the manuscript. However, Hu and Gao (2008, complete reference given below) estimated the average concentration of Li in the upper continental crust is 41 mg/kg. This is

higher than even the US study where the median Li content was about 20 mg/kg using a total extraction method. I think a brief discussion about reported Li concentration being totally dependent on the extraction used would be useful in the text.

Reference for Hu and Gao (2008):

Hu, Z., and Gao, S., 2008. Upper crustal abundances of trace elements—A revision and update. Chemical Geology 253 (3-4), 205–221.

No details on what specific Li were reported by Schrauzer. Hence, we removed them to avoid confusion. We have added reference from Teng et al., which analysed the Lithium concentration using HF+HNO_3 extractable.

Discussion about dependency of Li concentration on the extraction method used is beyond the scope of this paper because we did not measure the samples using various method to be able to document this.

14. Lines 250-254. The concentration ranges of the various sized circles in Figure 1 make it difficult for the reader to see where the higher concentrations of Li occur. I suggest that you show a range and median for the samples in each of the areas discussed in the text (i.e., Cape York, Goldfields-Esperance, etc.).

    We have added descriptive statistics showing the range and median concentration of Li in the areas mentioned.

15. Lines 55, 265, and 380. The authors cite "Foregs (2006)". However, there is no Foregs (2006) in the References. I think the authors have this listed as "Geochemical Atlas of Europe" in the References. The correct citation in the text should be "De Vos, Tarvainen, et al. (2006)." Then the complete reference should be as follows:

    De Vos, W., Tarvainen, T., Salminen, R., Reeder, S., De Vivo, B., Demetriades, A., Pirc, S., Batista, M.J., Marsina, K., Ottesen, R.T., O'Connor, P.J., Bidovec, M., Lima, A., Siewers, U., Smith, B., Taylor, H., Shaw, R., Salpeteur, I., Gregorauskiene, V., Halamic, J., Slaninka, I., Lax, K., Gravesen, P., Birke, M., Breward, N., Ander, E.L., Jordan, G., Duris, M., Klein, P., Locutura, J., Bel-lan, A., Pasieczna, A., Lis, J., Mazreku, A., Gilucis, A., Heitzmann, P., Klaver, G. and Petersell, V. (2006). *Geochemical Atlas of Europe. Part 2 – Interpretation of geochemical maps, Additional Tables, Figures, Maps and related publications*. Geological Survey of Finland, Espoo, Finland, p. 225-228. http://weppi.gtk.fi/publ/foregsatlas/text/Li.pdf

    We have updated the references accordingly.

16. Line 315. Here again the authors refer to “releveling” the NAGS data sets. As stated in comment #7, it would be useful to discuss this leveling process in a bit more detail.

    We have added more detail in section 2.1.

17. Line 423. Delete “with anomalous Li concentration” at the end of this sentence.

    We have removed the words “anomalous Li concentration.

essd-2022-418

Responses to Reviewer 3

This study aims to predict and map lithium (Li) concentration in soil across Australia using a digital soil mapping framework and environmental covariates. The model was developed using a Cubist regression tree algorithm and validated on an independent Northern Australia Geochemical Survey dataset, showing good prediction for the top depth. The importance of variables indicates that Landsat 30+ Barest Earth bands and gamma radiometric dose have a strong impact on Li prediction.

My overall impression is that despite the not relay convincing prediction power and out-of-sample verification that needs to be extended authors rigorously planned their work and did their best. For example, the set of statistics chosen to evaluate prediction performance was chosen wisely, the methodology seems appropriate (although I have questions about that), but the MS is quite worthy of being published in ESSD after the questions of all reviewers are answered. Here I underline that in my review I primarily evaluated the work from a methodological aspect.

We thank the reviewer for the constructive comments, and will address the questions accordingly.

Major comments

1. I am curious why the Cubist model was chosen. There is no literature review on which other approach could have been used in this particular exercise. Did you check other tree-based machine learning algorithms like Random forest for example. Since Random Forest can capture complex non-linear relationships between input variables and output variables by creating multiple decision trees and combining them, whereas Cubist uses linear models to estimate the output values for each leaf node of the decision trees, which may not be able to capture complex non-linear relationships.
   I work in isotope hydrology and before conducting the prediction of isoscapes we conducted through research on which approach would be most suitable for the task keeping in mind the number of predictors, the drivers of the parameter etc. Such a comparison would be necessary to be:
   - cited
   - conducted and places in supplement
   - or published in another study.

   See for example: https://doi.org/10.1016/j.jhydrol.2023.129129. In addition, a flowchart should also be added to the MS, possibly in supplement to help reproduce the steps taken.

We did not compare the performance of Cubist model to other machine learning models. Both Cubist and random forest utilise decision tree approaches. Furthermore, Cubist has been shown to show similar performance / outperform random forest; and the results are more interpretable compared to random forest.

Khaledian, Y. and Miller, B. A.: Selecting appropriate machine learning methods for digital soil mapping, Applied Mathematical Modelling, 81, 401-418, https://doi.org/10.1016/j.apm.2019.12.016, 2020.
Pouladi, N., Møller, A. B., Tabatabai, S., and Greve, M. H.: Mapping soil organic matter contents at field level with Cubist, Random Forest and kriging, Geoderma, 342, 85-92, https://doi.org/10.1016/j.geoderma.2019.02.019, 2019.

We have added a flowchart into the MS.

2. How was the preprocessing conducted, outlyers, extreme values? See for example the ultimate two paragraphs of Sect. 2.1 in https://doi.org/10.1016/j.jhydrol.2023.129129 . Did you check the outliers in the input data, I'm not sure how the Cubist model can handle them, as it uses linear models to estimate output values for each leaf node of the decision trees. Outliers can have a large impact on the estimated output values of the linear models, which can lead to inaccurate predictions.

   We did not remove any high extreme values. These values are used to identify areas with anomalous concentration of Li. The values below detection limit (<0.1) are replaced with half of the detection limit (0.05). Furthermore, NGSA dataset underwent a thorough data quality assessment, explained in:

   de Caritat, P. and Cooper, M.: National Geochemical Survey of Australia: Data Quality Assessment. Record 2011/021, 2011b.

3. A more detailed description of the used metrics is requires, since all of these account for different kind of errors. E.g. the Lin's CCC measures both the correlation and the bias between the measured and predicted values and it provides a measure of the strength of the linear relationship between the two value sets, while accounting for the magnitude of the differences.
   In addition, references should be inlcuded, e.g. Lin, 1989 https://www.jstor.org/stable/2532051

   We have updated the description of the used metric and added the relevance citation to Lin's CCC.

4. The argument in L412-413 is acceptable, but isn't there a possibility to validate the results with data from other regions or conduct a pilot study

from elsewhere? In a study in a journal as ESSD (upper 1 percentile in SJR) it would be expected to provide an even broader validation scheme, or give an extensive explanation on why this is not possible.

The trained model could not be used to validate data from other regions, as it was only trained using environmental covariates within Australia, in particular as the soil may be quite different here.

Minor comments

1. It might be more appropriate to categorize the predictors according to which ones were considered static (do not change over time) and which ones were considered dynamic (can change over time).

   In this case, we are following the scorpan concept, and hence the covariates were grouped following the scorpan method.

2. It was not discussed earlier, is a linear relationship (measured by Pearson r) required, or is there a nonlinear relationship expected between the predictors and Li content. Please elaborate on this.

   Linear relationship between the predictors and Li content are not necessarily needed. It is part of the exploratory data analysis. The machine learning model is used to develop the relationship between the predictors and Li content. The linear correlation can potentially be used to explain the usage of predictors within a model.

3. L262: What were these correlation values for Al, B, Fe..., a table should be included e.g. in the supplement.

   The correlation tables have been added in the supplementary materials.

4. The significance values should be reported and all the statistics in APA style. https://www.socscistatistics.com/tutorials/correlation/default.aspx

   We have added statistical significance for the Pearson's correlation (see Table SM1).

5. L407: This limitation is very important and must be mentioned in the abstract, in addition, it could even be incorporated into the title.

   We mentioned that our samples are collected from the catchment outlet floodplain sediments in abstract and within the Materials section.

Miscellaneous

1. Add spaces before and after mathematical operators.

   We have updated the text.

2. L20 and all other places use superscript for measurement units kg^{-1}

   We have updated the text.

3. Variables should be in italics.

   We have italised the parameters used to tune the model.

4. A more detailed description is needed on the boxplots explaining what is in the figure, see e.g. caption of Fig. 3 in https://doi.org/10.1016/j.jhydrol.2022.128925

   We have added more description in the caption.