Reply to Referee #1 Dr. Nathan Bowman's comments and suggestions

General comments:

Uncertainty quantification is a commonly studied aspect of mineral prospectivity mapping, and this manuscript presents a potentially valuable framework for individually quantifying different types of uncertainty. This is an approach that could be adopted by geologists in both academia and industry to guide and improve their own MPM studies. However, several key methodological steps regarding the treatment of geochemical and geophysical data are either insufficiently described or ambiguously presented, making it difficult to assess the validity of the uncertainty analysis and the robustness of the conclusions. Clarifying these aspects is essential, as they directly underpin the central claims of the study. Furthermore, the discussion section does not adequately consider relevant prior work, nor does it clearly position this study in relation to existing methods. Although the code used to produce the results is included as supplementary material, the data provided are insufficient to resolve the issues described above or to fully reproduce the authors' workflow. Many sections of the manuscript are repetitive and contain redundant information. The figures are of variable quality, and the figure captions lack sufficient description. With major revisions to improve the structure of the manuscript, clarify the methodology, and strengthen the discussion of data analysis and interpretation, this manuscript has the potential to make a meaningful contribution to the field.

Re: We sincerely appreciate your valuable and constructive comments. We have carefully considered all of the comments and revised the manuscript to address the concerns raised.

Regarding the treatment of geochemical and geophysical data, we have expanded and clarified the methodology section to clearly describe the preprocessing procedures used to construct the evidential layers. Regarding the discussion section, we have strengthened both the literature review and the discussion by incorporating additional relevant studies, and providing a comparison between our proposed framework and existing uncertainty quantification methods. Regarding code and data availability, we have enhanced the supplementary materials to provide additional details on data processing, model implementation, parameter settings, and uncertainty quantification procedures. In addition, we have carefully revised the manuscript structure and presentation by removing repetitive descriptions, reducing unnecessary redundancy, and improving the overall organization and readability. All figures have been improved to enhance their clarity and resolution, and the figure captions have been expanded to provide more comprehensive descriptions of the results and interpretation. We believe these revisions have substantially improved the manuscript, and we sincerely appreciate the reviewer's insightful suggestions, which have helped strengthen the quality of this study.

Specific comments:

While individual sections of the manuscript are well written, several sections could be combined and simplified due to repetitive content. In particular, Section 2.1 (lines 129-139) repeats information from earlier sections (lines 75-90), and much of the methodology section provides general background rather than describing the specific methods the authors used. The actual workflow is described later (Sections 4 and 5, lines 301-404). Sections 4 and 5 are very well written, very descriptive, and should be the information provided in the methodology. The authors are encouraged to streamline the methodology by reducing background material and consolidating relevant content into a single, coherent section.

Re: Thank you for your constructive suggestions. We have carefully streamlined the Methodology section by reducing general background information, removing redundant descriptions, and focusing on the methods implemented in this study. We have also reorganized the manuscript structure by relocating relevant descriptions of the proposed workflow from the Sections 4 to Section 2, where they are more appropriately presented as part of the methodology.

Many figure captions are vague and insufficiently descriptive. This often leads to confusion about what is being shown in the figures and how they relate to the workflow and results. Figure captions should be expanded to clearly explain what is being shown and their relevance to what is being discussed in the text. Figure 3 should be redrafted: It is not clear in either the text, the figure captions, or the illustration itself what is being shown or how it relates to the implementation of a Bayesian CNN. It is also unclear what is shown in Figure 16: do the mean vs uncertainty scatter plots represent pixel-wise values from the mean and variance maps? These points should be clarified in both the text and figure captions.

Re: Thank you for your suggestions. We have revised all figure captions to provide comprehensive descriptions of the displayed information.

In particular, Figure 3 has been redesigned to more clearly illustrate the architecture and implementation process of the Bayesian CNN. The accompanying text and figure caption have been revised to better to explain its role within the proposed framework. We have also expanded the captions and related descriptions for Figures 1, 2, and 16 to clarify their contents and interpretation. The revised captions are as follows:

Figure 1: Workflow of geological evidential layer simulation for data uncertainty modeling. Geological data are transformed into weighted evidential features using a distance-decay weighting function, and uncertainty is modeled by randomly sampling the maximum controlling distance to generate multiple simulation layers.

Figure 2: Workflow of geochemical and geophysical evidential layer simulation for data uncertainty modeling. Discrete observations are used as conditioning data, while the original gridded evidential layer serves as TI. For each unsampled node, DS algorithm sequentially scans the TI to find the best-matching local pattern and assigns its central value, generating multiple stochastic realizations that preserve spatial continuity and represent data uncertainty.

Figure 16: Scatter plots of mean probability versus uncertainty: (a) data uncertainty, (b) model uncertainty. The x-axis denotes the mean prediction probability obtained from multiple stochastic forward passes and the y-axis denotes the corresponding uncertainty (entropy).

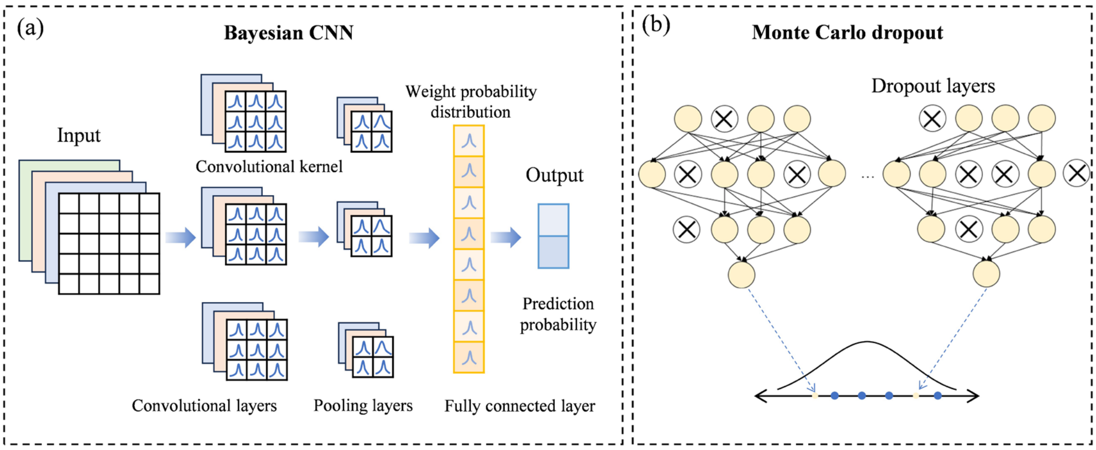

Figure 3: Workflow of model uncertainty modeling: (a) Bayesian CNN, (b) Monte Carlo dropout. Model uncertainty is quantified by treating network weights probabilistically in a Bayesian CNN. Bayesian inference is approximated using Monte Carlo dropout by activating dropout during inference and performing multiple stochastic forward passes to estimate uncertainty.

The inclusion of data and code in the appendices is appreciated and supports reproducibility. However, the provided datasets (two .tif files) are insufficient. In particular, the file evidence.tif is not georeferenced, and it is unclear what it represents.

Re: In the revised supplementary materials, we have uploaded all data and code related to the simulation and quantification of uncertainty, including:

(1) Geological, geochemical, and geophysical evidence layers used in mineral prospectivity mapping

(2) Original evidence layers and simulated evidence layers

(2) Code for continuous and discrete data uncertainty modeling

(3) Codes for data uncertainty, model uncertainty, and prediction uncertainty quantification

(4) Detailed documentation describing the complete workflow, required inputs, outputs, software environment, and execution procedures.

The complete repository link and permanent identifier (DOI) are provided below:

https://doi.org/10.5281/zenodo.20793560

Terminology relating to the types of uncertainty is inconsistent throughout the manuscript. Four types of uncertainty are initially defined: conceptual uncertainty, data uncertainty, variable uncertainty, and integration uncertainty (after Zuo et al., 2021; lines 55-60). The stated goal of the study is to quantify data uncertainty and "model" uncertainty (line 71). The authors then subsequently define data uncertainty as "limited sampling, observational errors, and the selection of negative samples, which is primarily expressed during the construction of evidential features, including ore-controlling geological structures, geochemical anomalies, and geophysical anomalies" (lines 72 to 74), which combines aspects of both data uncertainty and variable uncertainty. The authors then define model uncertainty as "the variability in deep learning architecture and parameters used to integrate these evidential features (lines 74-75)", which corresponds more closely to the earlier definition of integration uncertainty (line 59). This inconsistency persists throughout the manuscript and creates confusion, as it is unclear which types of uncertainty the different methods are intended to address. For example, on line 332, "parameter uncertainty" is used to describe what was previously defined as variable uncertainty. The authors are encouraged to adopt a clear terminology framework and apply it consistently throughout the manuscript.

Re: Thank you for this insightful comment. We have standardized the terminology related to uncertainty throughout the revised manuscript.

To avoid ambiguity, we clarify that this study focuses on the quantification of data-related uncertainty and model-related uncertainty in MPM. Data uncertainty considers limited spatial resolution of geochemical/geophysical data, subjective interpretations of geological model, and the selection of negative samples, which is primarily expressed during the construction of geological, geochemical, and geophysical evidential layers. Model uncertainty considers the variability in deep learning architecture and parameters used to integrate these evidential features. We have also revised the relevant sections by replacing ambiguous or inconsistent terms. These revisions provide a clearer distinction between the uncertainty sources addressed in this study and improve the consistency of the manuscript.

It is not clearly justified why MPS-DS is applied to geochemical and geophysical data. The authors state that MPS-DS is used to interpolate these datasets (explicitly stated for geochemical data in line 345 and implied for geophysical data in line 354), stating that generating multiple realizations allows them to quantify both data and model uncertainty (lines 364-365). However, other more commonly used methods of interpolation such as kriging also allow uncertainty to be quantified. Therefore, the authors should clarify why their approach is advantageous relative to established interpolation methods, particularly in relation to uncertainty quantification.

Re: Thank you for this valuable comment. To model the spatial distribution patterns of geochemical and geophysical data, the DS algorithm is employed to transform discrete geochemical measurements into spatially continuous evidential layers on a 500 m grid from the original 2 km scale. This enhancement in spatial resolution inevitably introduces data uncertainty associated with the estimation of unsampled locations. Although traditional interpolation methods such as kriging can provide estimation of uncertainty, they are primarily based on two-point statistics and typically rely on assumptions of stationarity and linear spatial correlation structures. As a result, they may not adequately capture complex, non-linear, and spatially heterogeneous patterns that commonly characterize geochemical and geophysical data. In contrast, MPS-DS utilizes multiple-point statistics and can reproduce higher-order spatial relationships contained within the training image or observed data patterns. Furthermore, MPS-DS generates multiple realizations that preserve spatial variability and geological continuity, allowing uncertainty to be propagated through subsequent modeling steps rather than being represented solely by local estimation variance. This capability makes MPS-DS particularly suitable for the uncertainty quantification framework proposed in this study.

There is no reference in the text about origin or source of the aeromagnetic data, and the treatment of this dataset raises several concerns. The authors suggest that uncertainty from arises interpolation due low sampling density and can lead to over-smoothed results (lines 83-86, 194-196, 583-585). However, aeromagnetic data typically consist of dense measurements collected along flight lines, with regular spacing of flight lines across the study area. Yet figure 8b shows equidistant data points with extensive coverage over the entire region, which does not appear consistent with how aeromagnetic data are usually collected, nor does it reflect the "limited sampling density" issue the authors wish to address. Additionally, it is unclear how the vertical first-order derivative (1VD) was applied to the aeromagnetic data (lines 295-296). It should be clarified whether:

1. the 1VD transformation was applied to raw survey data prior to simulation, or
2. the data were first gridded, then 1VD transformed, and subsequently used in simulation.

Given the regular spacing shown in Figure 8b, the latter seems more likely. In either case, the processing workflow should be clearly described in both the main text and the figure captions.

Re: Thank you for this valuable comment. The aeromagnetic data used in this study were obtained from the China Geological Survey at a 1:200,000 scale. Importantly, these data were provided as a gridded raster rather than raw flight-line measurements. Regarding the processing workflow, the 1VD was first calculated from the gridded aeromagnetic data, and the resulting 1VD raster was subsequently used as the input for DS simulation. Finally, the original 2000 m gridded aeromagnetic data were used as the training image to generate multiple stochastic realizations on a 500 m grid. Therefore, the uncertainty considered in this study is not associated with the original flight-line acquisition geometry, but with the enhancement of spatial resolution for archiving aeromagnetic evidence layer using DS simulation. We have revised the manuscript to explicitly describe this workflow in the Dataset and Evidential layers simulation section and updated the corresponding figure caption.

Similar concerns apply to the geochemical data. The authors state that "geochemical data provide direct evidence of elemental anomalies linked to gold mineralization" (lines 281-282), yet the data are stream sediment samples (line 288). These samples are not in situ and therefore do not provide direct evidence of the magmatic-hydrothermal style of gold mineralisation from the study area (line 273). Rather the data reflects a combination of processes, including erosion, transport, and mixing of multiple source lithologies. As such these samples likely reflect the aggregate geochemical composition of multiple outcrops sourced from multiple locations, not just mineralised outcrop. This distinction is not addressed in the manuscript.

Re: We agree that stream sediment geochemical data do not provide direct evidence of magmatic-hydrothermal gold mineralization. Instead, they represent integrated geochemical signatures resulting from weathering, erosion, sediment transport, and mixing of materials derived from multiple upstream lithologies. To avoid overstatement, we have revised the manuscript accordingly. We have also clarified that stream sediment geochemical anomalies represent the integrated effects of multiple geological and surficial processes and therefore serve as proxy indicators for mineral prospectivity rather than direct evidence of mineralization.

Furthermore, the stream sediment samples appear to be located on a regular grid (Figure 8a), which is atypical for stream sediment sampling. This suggests that the data shown in figure 8a is either a mix of outcrop, regolith, and stream sediment samples collected on a grid, or that they are stream sediment samples that have been interpolated prior to simulation. Again, the processing workflow should be clearly explained in both the text and figure captions.

Considering the issues discussed above (vague and undescriptive figure captions regarding geophysical and geochemical data, ambiguous descriptions of how the data was collected and treated, and the presentation of geochemical samples and geophysical data on apparently regular grids which is atypically for how the data is collected), the current presentation of the data gives the impression to the reader that these datasets were interpolated prior to MPS-DS. If both geochemical and geophysical datasets were interpolated prior to MPS-DS, this undermines the claim that MPS-DS “alleviates smoothing effects at unsampled locations” (lines 195-196), as it would instead be sampling from already smoothed data. If this is not the case, the authors should explicitly clarify what is represented in Figure 8 and provide a clear description in the text of how the geochemical and geophysical datasets were collected and processed prior to simulation to avoid confusion.

Re: Thank you for this thoughtful comment. We have revised the manuscript to clarify both the source of the geochemical data and the preprocessing workflow.

The stream sediment geochemical data were sourced from the Chinese National Geochemical Mapping Project at a 1:200,000 scale. Each sample contains concentrations of 39 major and trace elements. In this regional geochemical survey, stream sediment samples were collected at an average density of one sample per 1 km², and four adjacent samples were composited to produce one representative sample for a 4 km²area grid cell. Consequently, the final geochemical dataset exhibits an approximately regular spatial grid, as shown in Figure 8a. Geochemical data are typical compositional data, which can induce false spatial and statistical correlations among elements. Prior to DS simulation, the geochemical data were preprocessed using the centered log-ratio transformation to eliminate the closure effect inherent in compositional data. The transformed geochemical maps were subsequently used as the input for DS simulation. We have revised the manuscript and the figure caption.

While the quantification of uncertainty in MPM (and geological models in general) is an active area of research, the discussion section contains only one reference, which is to support the use of ROC-AUC (line 532). There is no mention or discussion of similar studies, or how this study either compliments or improves upon the methods and findings of previous research. Furthermore, there are two claims made in the discussion and conclusion that are not supported by the data and results presented:

1. The authors claim that uncertainty quantification "reveals the key geological, geochemical, and geophysical factors that most strongly control mineralization, offering a deep understanding of the underlying ore-forming processes" (lines 551-552). However, only geological factors are briefly discussed (that the faults and Xiong'er rock group is more prospective than the Yanshanian intrusions, lines 439-443, 463). Geochemical and geophysical factors are not mentioned. The current discussion does not highlight how the methods reveal a "deep understanding" of ore forming processes

2. The authors state that "data uncertainty in MPM arises from limited sampling density of geochemical and geophysical data, as well as subjective interpretation in geological modelling…" (lines 583-584). As discussed above, the geochemical and geophysical datasets as presented in the text and figures appear extensive and do not show clear evidence of limited sampling density. As written, the manuscript does not adequately support or address this conclusion.

Re: Thank you for this constructive comment. We have revised the Discussion section by incorporating comparisons with previous studies and highlighting the novelty of the proposed framework.

Uncertainty quantification has become an increasingly important component of deep learning-based MPM. Currently, existing studies primarily focused on uncertainty associated with parameters optimization, network structures, or prediction models selection using approaches such as Bayesian neural networks and ensemble learning (Daviran et al., 2025; Hoseinzade et al., 2025; Nwaila et al., 2026; Liu, 2025, 2026). Although these methods provide valuable measures of prediction confidence, they generally assume that the input evidential layers are deterministic. In practice, the construction of geological, geochemical, and geophysical evidential layers is subject to substantial uncertainty. Neglecting these sources of uncertainty may result in unreliable prospectivity predictions. The proposed framework in this study addresses this limitation by integrating uncertainty quantification for both the input evidential layers and the prediction model. By transforming uncertainty from the evidential layers to the final prospectivity predictions, the proposed approach provides a more comprehensive characterization of prediction uncertainty throughout the entire MPM workflow than approaches relying solely on model uncertainty estimation.

1. In the metallogenic system, geological features usually offer direct evidence, while aeromagnetic data and geochemical data provide indirect evidence of regional geophysical and elemental anomalies linked to mineralization. In the present study, the discussion mainly focuses on the geological controls represented by faults and formations, while the roles of individual geochemical and geophysical evidential layers were not analyzed separately. Accordingly, we have revised the manuscript to avoid overinterpretation. The revised sentence is:

"Uncertainty quantification provides essential insights into the reliability of MPM. On the one hand, it delineates regions characterized by low uncertainty and high prospectivity probability, thereby guiding exploration efforts toward areas with the greatest potential of new mineral discoveries. On the other hand, it facilitates the interpretation of spatial relationships between mineral prospectivity and major geological features."

2. We agree that the previous wording did not adequately explain the source of data uncertainty associated with the geochemical and geophysical datasets. The original geochemical and geophysical datasets were acquired at a scale of 1:200,000, corresponding to a spatial resolution of approximately 2 km. However, the MPM in this study was performed on a 500 m grid. This enhancement in spatial resolution inevitably introduces data uncertainty associated with the estimation of unsampled locations. To address this scale mismatch, DS was employed to generate multiple conditional realizations on the 500 m grid. Rather than eliminating uncertainty caused by sparse sampling, DS models and propagates this uncertainty by infilling unsampled locations and enhancing the spatial resolution while preserving the spatial structures contained in the original datasets. We have revised the manuscript to clarify that

data uncertainty, especially geochemical and geophysical data, arises primarily from the coarse spatial resolution of the original regional datasets relative to the finer modeling grid.

Reference:

Daviran, M., Ghezelbash, R., Hajihosseinlou, M., and Maghsoudi, A.: Uncertainty quantification in genetic algorithm-optimized artificial intelligence-based mineral prospectivity models: Automated hyperparameter tuning for support vector machines and random forest, Modeling Earth Systems and Environment, 11(1), 10, https://doi.org/10.1007/s40808-024-02176-z, 2025.

Hoseinzade, Z., Shojaei, M., Khademi, F., Mokhtari, A. R., and Saremi, M.: Integration of deep learning models for mineral prospectivity mapping: A novel Bayesian index approach to reducing uncertainty in exploration, Modeling Earth Systems and Environment, 11(3), 161, https://doi.org/10.1007/s40808-025-02342-x, 2025.

Nwaila, G. T., Durrheim, R. J., Frimmel, H. E., and Carranza, E. J. M.: Robust uncertainty analysis in mineral prospectivity mapping: A prototype for Fe–Mn exploration in South Africa, Natural Resources Research, 35(2), 901–931, https://doi.org/10.1007/s11053-025-10615-6, 2026.

Liu, Y.: A spatially aware evidential deep learning framework for mineral prospectivity mapping and uncertainty evaluation, Mathematical Geosciences, https://doi.org/10.1007/s11004-025-10266-6, 2026.

Liu, Y.: Quantifying uncertainty of mineral prediction using a novel Bayesian deep learning framework, Artificial Intelligence in Geosciences, 6(2), 100164, https://doi.org/10.1016/j.aiig.2025.100164, 2025.

Line 78: The statement that weights-of-evidence uses "binary weights (0 or 1)" is inaccurate. The weightings used in WoE are typically logged likelihood ratios. The authors are referred to Bonham-Carter (1994) for a discussion of the method.

Re: Thanks, we have corrected this inaccurate description in the revised manuscript. The revised sentence is:

"Traditional approaches usually assign binary weights (0 or 1) based on the presence or absence of mineral deposits, thereby introducing uncertainty due to subjective interpretations of mineralization processes. Improved approaches like fuzzy logic and fuzzy weights of evidence methods overcome this limitation by assigning continuous weights ranging from 0 to 1, which more realistically represent geological facts (Cheng and Agterberg, 1999; Lisitsin et al., 2014)."

Line 129-130: The statement that mineralization is controlled by "geological, geochemical, and geophysical features" is conceptually misleading. Mineralization is controlled by geological processes; geochemical and geophysical data provide evidence for these processes. The authors are referred to Hronsky & Kreuzer (2019) and McCuaig et al. (2010) for a discussion on the distinction of mineral system processes and the use of evidential maps in mineral targeting.

Re: Thanks, we have corrected this inaccurate description and cited the two references. The revised sentence is:

"Mineralization is controlled by the combined influence of multiple geological processes, and geochemical and geophysical data provide indirect evidence for these processes (McCuaig et al., 2010; Hronsky and Kreuzer, 2019). For example, faults often serve as pathways for hydrothermal fluid migration, intrusions supply heat sources and mineralizing materials, and geological formations create favorable conditions for mineral precipitation. Geochemical anomalies reflect the dispersion and enrichment of ore-related elements, whereas geophysical signatures reveal subsurface structures and physical property contrasts associated with mineralization."

Lines 204-205: The use of randomly selected background samples as negative data is an appropriate and widely used technique in multiple disciples, not just geoscience. Additional references could be included to strengthen this section, for example Montsion et al. (2024), as well as literature from spatial ecology on pseudo-absence data (e.g. Wisz & Guisan, 2009; Barbet-Massin et al., 2012). There are other methods to generate negative training data for MPM, such as using drill hole assay data depleted in a commodity of interest, or the location of mineral systems that are geologically different from the system of interest that may also be mentioned (see Lindsay et al. 2022).

Re: Thanks, we have revised this part and cited related references in the revised manuscript. The revised sentence is:

"Additionally, data-driven MPM generally relies on both positive and negative training samples to establish the relationship between predictor variables and mineralization. Positive samples are typically derived from known mineral deposit locations. In contrast, negative samples are more challenging to define because the absence of known mineralization does not necessarily indicate the absence of undiscovered deposits. Consequently, negative samples are commonly constructed from randomly selected background points under geological constraints, known occurrences of other mineral deposit types, or drill holes with economically insignificant mineralization (Carranza et al., 2008; Nykänen et al., 2015; Lindsay et al., 2022; Montsion et al., 2024). Previous research has demonstrated that the selection strategy for negative samples can substantially affect the predictive performance of MPM (Zuo and Wang, 2020). Therefore, developing reliable strategies for constructing negative samples remains a critical challenge in MPM."

References:

Carranza, E., Hale, M., and Faassen, C.: Selection of coherent deposit-type locations and their application in data-driven mineral prospectivity mapping, Ore Geology Reviews, 33(3–4), 536–558, https://doi.org/10.1016/j.oregeorev.2007.07.001, 2008.

Nykänen, V., Lahti, I., Niiranen, T., and Korhonen, K.: Receiver operating characteristics (ROC) as validation tool for prospectivity models—A magmatic Ni–Cu case study from the Central Lapland Greenstone Belt, Northern Finland, Ore Geology Reviews, 71, 853–860, https://doi.org/10.1016/j.oregeorev.2014.09.007, 2015.

Lindsay, M. D., Piechocka, A. M., Jessell, M. W., Scalzo, R., Giraud, J., Pirot, G., and Cripps, E.: Assessing the impact of conceptual mineral systems uncertainty on prospectivity predictions, Geoscience Frontiers, 13(6), 101435, https://doi.org/10.1016/j.gsf.2022.101435, 2022.

Montsion, R. M., Perrouty, S., Lindsay, M. D., Jessell, M. W., and Sherlock, R.: Development and application of feature engineered geological layers for ranking magmatic, volcanogenic, and orogenic system components in Archean greenstone belts, Geoscience Frontiers, 15(2), 101759, https://doi.org/10.1016/j.gsf.2023.101759, 2024.

Zuo, R., and Wang, Z.: Effects of random negative training samples on mineral prospectivity mapping, Natural Resources Research, 29(6), 3443–3455, https://doi.org/10.1007/s11053-020-09706-3, 2020.

Line 358: The claim that Figure 9 "clearly demonstrates" that DS reproduces the spatial structure of geochemical and geophysical data is not adequately supported. A cumulative frequency distribution alone does not capture spatial structure, and the same cumulative frequency distribution could be reproduced by random data. More appropriate methods would include capture efficiency curves, Ripley's K function, or nearest-neighbour G function.

Re: Thanks for your valuable suggestion. We agree that cumulative frequency distribution curves alone only evaluate the statistical distribution of the simulated data and do not adequately characterize their spatial structure.

To address this issue, we introduced experimental variograms to quantitatively assess the similarity of the spatial structures between the observed and simulated patterns. Specifically, the cumulative frequency distribution curves are used to evaluate the preservation of the statistical distribution of the geochemical values, whereas the experimental variograms are used to assess the preservation of spatial continuity and autocorrelation. The cumulative frequency distribution curve (Figure 9a) clearly indicates that the DS reproduces the statistical distribution pattern of geochemical features in the study area. Furthermore, the close agreement between the variogram curves of the observed and simulated realizations (Figure 9b) indicates that the DS approach successfully reproduces the major spatial structures of the original geochemical layer. These results demonstrate that the simulations preserve not only the statistical distribution of the evidential layers but also their essential spatial dependence, enabling the generation of realistic geochemical evidence layers while simultaneously accounting for uncertainty associated with unsampled regions.

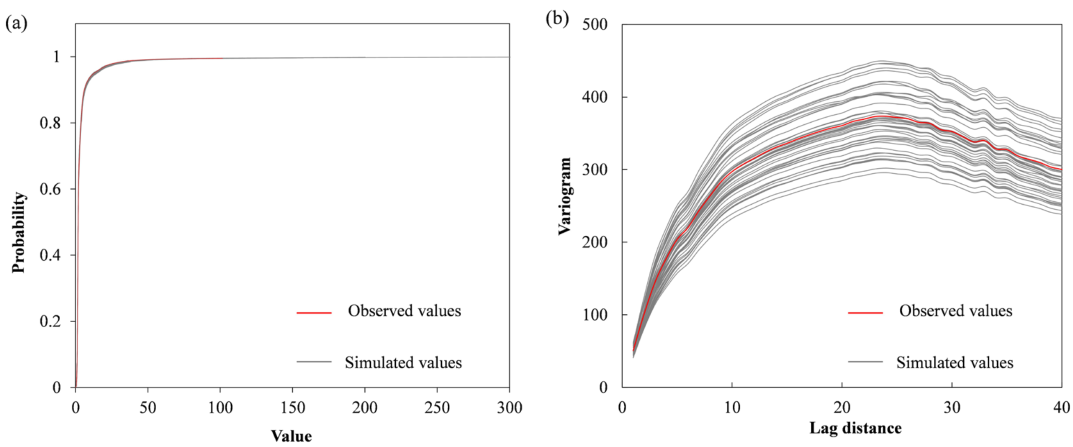

Figure 9: Cumulative frequency distribution curves (a) and experimental variogram curves (b) derived from 100 simulated Au geochemical layers.

Lines 432-435 state that, for data uncertainty quantification, “a deterministic CNN model is trained on simulated evidence layers to produce 100 prospectivity predictions.” I assume that these correspond to the 100 geological evidential layers produced through Monte Carlo sampling of maximum controlling distances, and the 100 geochemical and geophysical maps produced by MPS-DS. If this is the case, the authors state on lines 323-324, and again on lines 364-365, that multiple realizations of the evidential layers allow both the influence of data uncertainty and model uncertainty to be quantified. If the methods used to create these evidential maps incorporate both data uncertainty and model uncertainty, it is unclear in the text how they are subsequently used to quantify only data uncertainty.

Re: We apologize for the ambiguous expressions. In this context, multiple realizations of the geological, geochemical, and geophysical evidential layers are only used to evaluate data-related uncertainty. We have corrected this inaccurate description in the revised manuscript.

Technical corrections:

- Lines 12-14: the authors state the study proposes a framework to jointly evaluate data, model, and prediction uncertainty. However, on line 71 the authors state the goal is to quantify on model and data uncertainty. Either the introduction or the abstract needs to be updated to reflect the goals and aims of the study.

Re: We have carefully reviewed and revised both the Abstract and Introduction to ensure consistency in the description of the study objectives and uncertainty framework.

- Line 131: The term “formations” is ambiguous and should be clarified.

Re: Revised. The term “formations” is replaced by “geological formations”.

- Line 144: Replace “perspective” with “framework.”

Re: Revised.

- Line 178: Replace “prominence” with “prominent”

Re: Revised.

- Line 180: Remove “Theoretically”

Re: Revised.

- Line 192: The type of uncertainty referenced in this sentence is unclear and should be explicitly stated.

Re: The revised sentence is:

"By iteratively applying this procedure, DS generates multiple realizations, each honoring the conditioning data but differing in local spatial details due to stochastic sampling. This process transforms discrete samples into gridded maps that, on the one hand, alleviate smoothing effects at unsampled locations and, on the other hand, incorporate uncertainty arising from data gaps and limited spatial resolution."

- Line 306: Replace "quantitatively characterizing" with "quantifying".

Re: Revised.

- Line 378: Replace "are list in Table 2" with "are listed in Table 2"

Re: Revised.

- Table 2 is a repeat of Table 1. Likely an error in copying Table 2 to the manuscript.

Re: Revised. Table 2 summaries the hyperparameter settings of the Bayesian CNN.

Table 2: Hyperparameter settings of the Bayesian CNN.

Parameters	CNN
Window size	9
Dropout rate	0.3
Epoch	250
Batch size	32
Learning rate	0.00005
Optimizer	Adam
Loss function	Binary Cross Entropy
Activation function	Sigmoid

- Line 408-409: “Case studies” is not an appropriate term; consider using the terms “baseline model” and “experimental models.”

Re: Revised. The “Case studies” is replaced as “experimental models”.

- Line 490: Replace “jointly consideration” with “joint consideration”.

Re: Revised.

- Line 564: Remove “ultimately.”

Re: Revised.

Reply to Referee #2's comments and suggestions

In is paper, the authors present a method to propagate data and model uncertainty in a Mineral Prospectivity Mapping workflow. The manuscript is rather well structured and pleasant to read. However, some strong reformulation in the introduction, additional justifications and discussion point are missing. It would make the manuscript stronger in terms of addressed scope and impact. More details are given below.

Re: Thanks for your interests. We have carefully considered all of your suggestions and revised the manuscript accordingly. We sincerely appreciate your thoughtful comments, which have helped improve the quality of the manuscript.

Introduction

Lines 41-42: spread the citations across the specific mentioned techniques?

Re: We have revised the manuscript by distributing the citations to the corresponding techniques.

Lines 62-63: in the context of MPM, or in a more general context (e.g. modelling). These two categories are more general, than the four introduced earlier line 55. I would thus reverse the order of presenting these ways of categorizing uncertainties, from a more general approach (2 categories) to the more specific MPM context (4 categories).

Re: Thanks for your valuable suggestion. We have revised the structure of the Introduction in this revision.

Lines 71-115: this part of the introduction is hard to follow especially if the reader is not familiar with deep learning model for MPM. It would be helpful if the authors could recall the main components considered in MPM (e.g. structural features, degree of mineralization, etc.), the common architecture of such workflows (step 1-inferring layers, step 2- extracting features /, step 3- combining information to produce a map) and which methods are usually involved at each step. Then explain how uncertainty is currently ignored or considered at each step and what methods are available for it. Then clearly summarize the knowledge-gap or current limitations and define what you specifically address in this paper. Given the relatively important number of self-citations, it is important to clearly put in perspective the contribution and limitations of previous work by the authors to underline the novelty of this approach.

Re: Thanks for this insightful and constructive suggestion. We have reorganized this section of the Introduction for better expression. Specifically, we first introduce the concept of MPM and summarize the evolution of MPM methodologies, ranging from traditional statistical approaches to deep learning algorithms. We then describe the general workflow of deep learning-based MPM and emphasize that many existing deep approaches primarily focus on improving predictive performance while overlooking inherent sources of uncertainty, which may result in overconfident predictions. To address this issue, we review several classical definitions, sources, and classification schemes of uncertainty in MPM. Based on these concepts, we clarify that uncertainty in this study is categorized into data-related uncertainty and model-related uncertainty, and focus on the quantification of data uncertainty, model uncertainty, and their combined effect on total prediction uncertainty in MPM. Subsequently, we systematically review the origins, modeling approaches, recent advances, and limitations of data uncertainty and model uncertainty quantification in MPM. This discussion provides the motivation for developing the proposed uncertainty quantification framework.

2.1.1 Line 143: can you explain why multifractal singularity theory offers a rigorous mathematical foundation to simulate geological evidential layer uncertainty?

Re: Geological evidential layers are commonly constructed by discretizing geological features and transforming them into weighted evidential features using weighting function. This process introduces uncertainty because the weighting functions are typically determined based on geological knowledge and expert judgment rather than objective observations. To simulate the uncertainty associated with this weighting functions, multifractal singularity theory offers a mathematical foundation. However, multifractal singularity theory itself does not directly simulate uncertainty. Instead, it provides the basis for constructing weighting functions by describing the power-law relationship between mineralization intensity and the distance to geological features. The uncertainty arises because one of the key parameters in this weighting function, namely the maximum controlling distance (d_m), cannot be determined uniquely from observations. It is typically estimated empirically according to metallogenic models and geological expert judgment. Different reasonable choices of d_m lead to different weighting functions and, consequently, different geological evidential layers. To characterize this parameter uncertainty, we treat d_m as a stochastic variable and randomly sample it within a predefined interval to generate multiple realizations of the geological evidential layers. Therefore, multifractal singularity theory provides the physically and mathematically justified weighting model, while the stochastic sampling of d_m propagates the uncertainty associated with expert-defined parameters into the evidential layers.

The revised sentence is:

“The fractal dimension has proven effective for quantifying the spatial association between mineral deposits and geological features:

$\rho = c\varepsilon^{-D}$, (1)

where D denotes the fractal dimension and c is a constant. The fractal dimension has proven effective for quantifying the spatial association between mineral deposits and geological features, thereby facilitating the construction of weighting functions for various geological evidential features with the help of GIS. Generally, for a given location (i, j) near a geological feature, areas closer to the feature tend to contain more significant mineralization information and should therefore be assigned higher weights. Conversely, areas farther away are less relevant, or even irrelevant to mineralization, should be assigned lower weights. Accordingly, the weighting function for an evidential feature can be expressed as a piecewise formula:

$\omega_{ij} = \begin{cases} cd^{-D} & d \leq d_m \\ 0 & d > d_m \end{cases}$, (2)

where d denotes the distance from a given location (i, j) to the geological feature, d_m represents the maximum effective controlling distances of that feature. The weight ω quantifies the relative contribution of the geological feature to mineralization, with a larger ω indicating stronger ore-controlling effect.

In practice, the weight ω can be estimated by fitting a linear relationship to the log–log plot of ρ versus d using the least-squares method:

$\log(\rho) = -\omega \log(d) + c.$ (3)

In the formula 2, the controlling distance d_m, also known as the spatial extent of a feature’s influence, is typically set empirically based on metallogenic models, which vary according to expert interpretation. This subjectivity inevitably introduces data uncertainty into MPM. To account for this uncertainty, the parameter d_m is treated as a stochastic variable and randomly sampled within a predefined interval to generate multiple realizations of geological evidential features. Consequently, the weighting function for geological features is expressed as:

$$\omega = \frac{1}{1-{d_m}^{-D}}(x^{-D} - d_m), d_m \in [0, 1]. (4)$$

This formulation accounts for both the relative influence of geological features on mineralization and the uncertainty in the modeling process of evidential feature. The resulting weight values range continuously from 1 and 0. A weight of 1 indicates a strong statistical correlation between the geological feature and known mineral deposits, whereas decreasing values represent weaker spatial association. When the location x exceeds the maximum controlling distances d_m, the weight reduces to 0, implying that the geological feature in those regions exerts no measurable impact on mineralization.

After obtaining the weight ω, the geological evidence layer $g_{(\omega)}$ can be constructed as:

$$g_{(\omega)} = g \cdot \omega, (5)$$

where g is the original gridded geological feature and $g_{(\omega)}$ is the weighted geological evidence layer."

2.1.2 What training images are available for geochemical and geophysical evidential layers? In the example given in Figure 2, the TI seems to be a realization from a multi-Gaussian Random Field. What is the advantage of using the Direct Sampling algorithm in such a case? Would Sequential Gaussian Simulations not be sufficient?

Re: In this study, the training image (TI) is a raster map obtained from geochemical sampling points. In other words, the observed geochemical (or geophysical) measurements themselves are used as both the conditioning data and the TI. The primary reason for adopting DS rather than Sequential Gaussian Simulation (SGS) is that DS does not require Gaussian transformation or variogram model. Although SGS is an effective geostatistical simulation technique, it primarily reproduces two-point spatial statistics represented by the variogram and may smooth localized anomalies, thereby fails to reproduce more complex spatial configurations. In contrast, DS is a multiple-point simulation method that reproduces complex spatial patterns directly from the TI through neighborhood matching. Consequently, it can better preserve local spatial structures, nonlinear spatial relationships, and high-value anomalies that are important for MPM. Our objective is not only to interpolate unsampled locations but also to generate multiple realizations that preserve the observed spatial patterns while quantifying uncertainty. Therefore, DS was selected because it better satisfies these requirements without imposing restrictive distributional assumptions.

2.2 Using only a distance criterion to randomly sample negative samples seems a bit limited. Isn't there a risk of creating a bias? Would it be possible to consider other factors? Which ones? A map showing the locations of negative samples (or a sampling density map) would be meaningful to this study to compare later with the results.

Re: Thank you for this valuable comment. In general, negative samples are challenging to define because the absence of known mineralization does not necessarily indicate the absence of undiscovered deposits. Consequently, negative samples are commonly constructed from randomly selected background points under geological constraints, known occurrences of other mineral deposit types, or drill holes with economically insignificant mineralization (Nykänen et al., 2015; Lindsay et al., 2022; Montsion et al., 2024). In this study, we adopted a distance-based random sampling strategy in which negative samples were selected from areas sufficiently distant from known mineral deposits. Carranza et al. (2008) has demonstrated that randomly sampling points from regions located far from known deposits is an available strategy for generating negative samples in MPM uncertainty quantification. Nonetheless, we acknowledge that relying solely on a distance criterion may introduce sampling bias. We have added this discussion to the revised manuscript as a limitation of the current study.

References:

Carranza, E., Hale, M., and Faassen, C.: Selection of coherent deposit-type locations and their application in data-driven mineral prospectivity mapping, Ore Geology Reviews, 33(3–4), 536–558, https://doi.org/10.1016/j.oregeorev.2007.07.001, 2008.

Nykänen, V., Lahti, I., Niiranen, T., and Korhonen, K.: Receiver operating characteristics (ROC) as validation tool for prospectivity models—A magmatic Ni–Cu case study from the Central Lapland Greenstone Belt, Northern Finland, Ore Geology Reviews, 71, 853–860, https://doi.org/10.1016/j.oregeorev.2014.09.007, 2015.

Lindsay, M. D., Piechocka, A. M., Jessell, M. W., Scalzo, R., Giraud, J., Pirot, G., and Cripps, E.: Assessing the impact of conceptual mineral systems uncertainty on prospectivity predictions, Geoscience Frontiers, 13(6), 101435, https://doi.org/10.1016/j.gsf.2022.101435, 2022.

Montsion, R. M., Perrouty, S., Lindsay, M. D., Jessell, M. W., and Sherlock, R.: Development and application of feature engineered geological layers for ranking magmatic, volcanogenic, and orogenic system components in Archean greenstone belts, Geoscience Frontiers, 15(2), 101759, https://doi.org/10.1016/j.gsf.2023.101759, 2024.

Figure 7: it is not trivial to see the differences between the realizations from one row to the other. A fifth row displaying the standard deviation over the generated ensemble or weighted evidence layer might highlight the variability. A colour bar with the value range is missing for the evidence layers.

Re: Thank you for this valuable suggestion. Figure 7 has been revised to improve its clarity and interpretability. We have added a fifth row showing the standard deviation of generated geological evidential layers, which more clearly illustrates the spatial variability among the realizations. In addition, we have added color bars to the evidential layer maps.

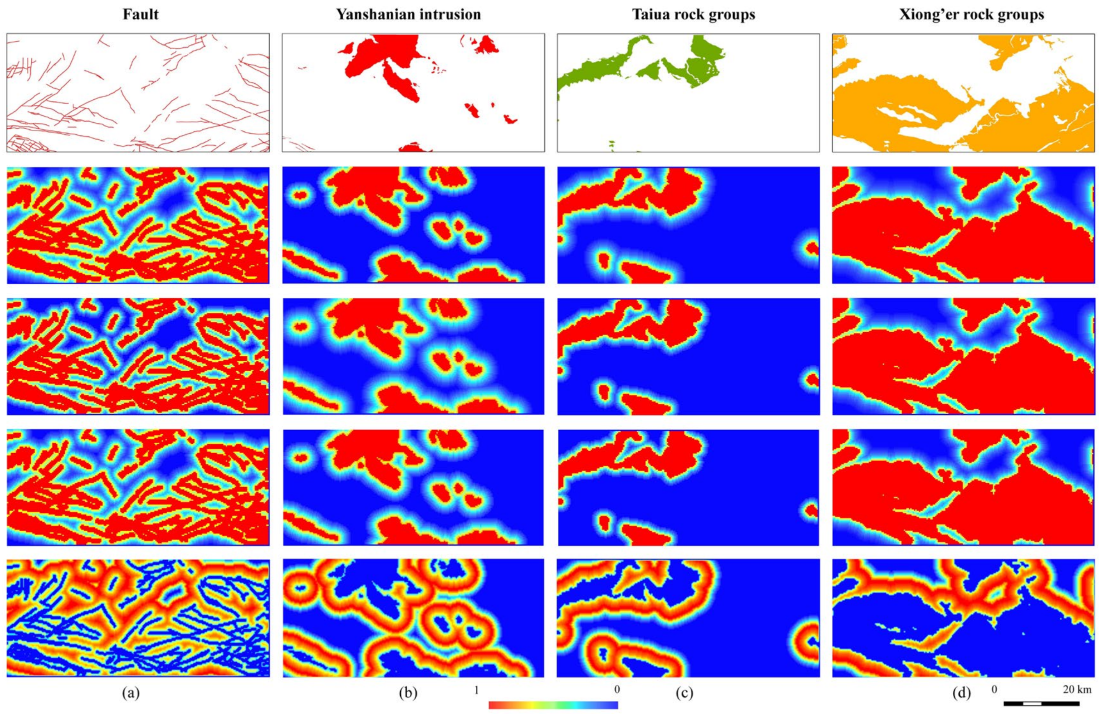

Figure 7: Weighted evidence layers: (a) faults, (b) Yanshanian igneous intrusions, (c) Taihua rock groups strata, (d) Xiong’ er rock groups strata. The first row is the original geological features, the second to fourth rows are the weighted evidence layers. The fifth row is the standard deviation of the generated evidence layer, which clearly illustrates the spatial variability among the realizations.

4.2 The Training Images need to be presented and clearly justified (what support the interpretation of such property field structures).

Re: Following previous applications of DS to geoscientific data, the observed geochemical and geophysical datasets are transformed into raster and used as the training images. To justify the suitability of these training images, we compared the simulated and observed evidential layers using cumulative frequency distributions and experimental variograms. Using Au as an example, the cumulative frequency distribution curve (Figure 9a) clearly demonstrate that the DS reproduces the statistical distribution pattern of geochemical features in the study area. The close agreement between the variogram curves of the observed and simulated realizations (Figure 9b) further indicates that the DS approach successfully reproduces the major spatial structures of the original geochemical layer. These results indicate that the selected training images adequately capture the spatial patterns required for stochastic simulation.

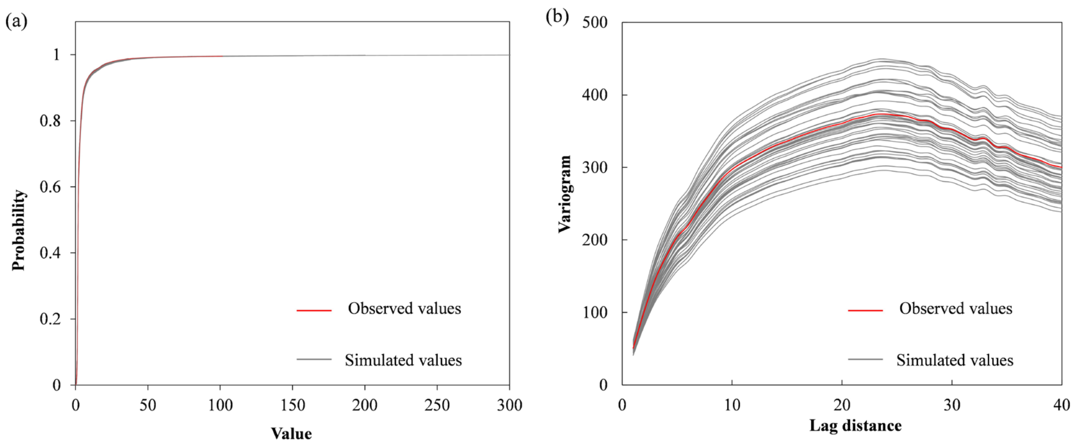

Figure 9: Cumulative frequency distribution curves (a) and experimental variogram curves (b) derived from 100 simulated Au geochemical layers.

Figure 8: though the differences are more visible than in Figure 7, a fifth row displaying the standard deviation of the simulated evidence layers would be great.

Re: Thank you for this valuable suggestion. We have added a fifth row showing the standard deviation of generated geochemical and geophysical evidential layers.

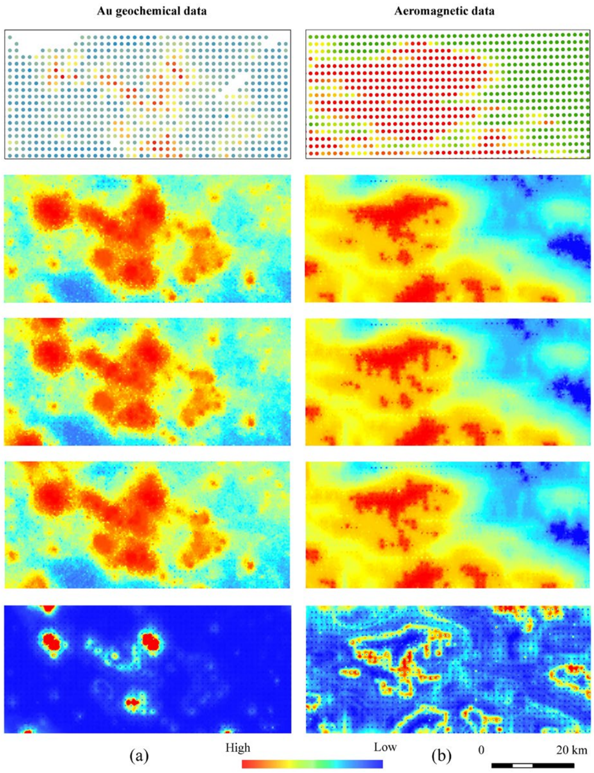

Figure 8: Simulated evidence layers: (a) Au geochemical layer, (b) geophysical aeromagnetic layer. The first row is the original geological features, the second to fourth rows are the simulated evidence layers. The fifth row is the standard deviation of the simulated evidence layer, which illustrates the spatial variability among the realizations.

5.1: how were the key hyper-parameters determined? How do you justify the values employed?

Re: The key hyperparameters are adjusted through grid search strategy. The optimal values are selected based on the best predictive performance and convergence curves of training loss and accuracy across epochs. We have added a description of the hyperparameter tuning in the revised manuscript.

Section 6

Line 405: 'Results' instead of 'Result'

Re: Revised.

Lines 452-453 "Overall, areas exhibiting high mean and low entropy values represent the most reliable exploration targets": Could you provide such an exploration map e.g. by multiplying the mean map with (1-entropy map)?

Re: Thanks for this valuable suggestion. We have added an exploration reliability map by combining the normalized mean prospectivity map with the entropy map.

To facilitate exploration target prioritization, an exploration reliability map (Figure 18) is delineated by integrating both the mean prediction probability with predictive uncertainty:

$R(p) = \bar{p}^* - H(p)^*$,

where $\bar{p}^*$ and $H(p)^*$ denote the normalized mean predictive prospectivity and normalized predictive entropy, respectively. This normalization places both variables on the same scale, allowing prediction potential and prediction reliability to be considered simultaneously. Areas with high values represent locations correspond to both high prospectivity and low uncertainty, and are therefore regarded as the most reliable exploration targets. Compared with the mean prospectivity map alone, this map suppresses regions with high prediction uncertainty while preserving areas with consistently high prospectivity, thereby providing a more reliable basis for exploration decision-making.

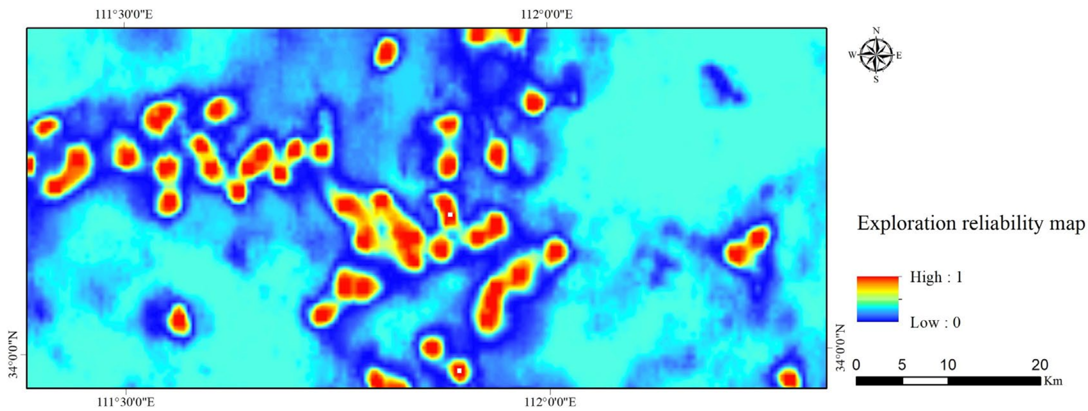

Figure 18: The exploration reliability map.

A single figure grouping figures 12, 13 and 14 might facilitate the comparison of the results.

Re: Thank you for the helpful suggestion. We carefully considered combining Figures 12-14 into a single composite figure. However, we decided to retain them as separate figures for two reasons. First, the three figures present the results of data uncertainty quantification, model uncertainty quantification, and prediction uncertainty quantification, respectively, which correspond to different sections of the Results and are discussed independently. Maintaining separate figures allows each uncertainty component to be presented and interpreted more clearly within its corresponding context. Second, each figure contains multiple high-resolution spatial maps. Combining them into a single figure would substantially increase the number of subfigures, reduce the display size of each panel, and compromise the readability of important spatial patterns and uncertainty distributions. Nevertheless, following the reviewer's suggestion, we have strengthened the connections among Figures 12-14 in the revised manuscript.

6.3 While 100 evidence layers realizations are used to quantify data uncertainty, how many models are generated using the Bayesian CNN for model uncertainty quantification? 100? from a single (original) set of evidence layers?

Re: Yes. This process is equivalent to sampling 100 subnetworks from a trained Bayesian CNN for model uncertainty quantification. Specifically, during inference, dropout layers were kept active, and 100 stochastic forward passes were performed on the same set of original evidence layers. Each forward pass samples a different thinned subnetwork by randomly deactivating neurons, which can be interpreted as sampling from the posterior distribution over the network parameters. The resulting ensemble of 100 predictions approximates the posterior predictive distribution, from which the predictive mean, variance, and entropy were calculated to quantify model uncertainty.

6.4 Where do the 10,000 realizations come from? 100 stochastic sets of evidence layers x 100 Bayesian CNN models?

Re: The 10,000 predictions do not arise from training 100 independent Bayesian CNN models with 100 stochastic sets of evidence layers. Instead, they are obtained by combining the stochastic sampling of both data and model uncertainties. Specifically, 100 stochastic realizations of the evidence layers are first generated to represent data uncertainty. For each realization, the same trained Bayesian CNN is evaluated using 100 Monte Carlo Dropout stochastic forward passes, resulting a total of $100 \times 100 = 10{,}000$ predictions.

7.1, the uncertainty decomposition should also be compared and discussed with figures 12 and 13. What are the similarities or differences? Why is it valuable to perform this decomposition rather than relying on data uncertainty and model uncertainty as generated in Figures 12 and 13?

Re: We have expanded the discussion to compare the uncertainty decomposition results (Figure 15) with the uncertainty maps shown in Figures 12 and 13.

Uncertainty decomposition analysis enables the partitioning of total predictive variance into components attributable to data variability and model stochasticity, thereby providing a clearer understanding of uncertainty sources and supporting more informed exploration decisions. Figures 12 and 13 characterize the spatial distributions of data uncertainty and model uncertainty, respectively. In contrast, uncertainty decomposition analysis does not represent uncertainty magnitude itself but decomposes the total prediction uncertainty into the relative contributions of data uncertainty and model uncertainty. This contribution map additionally reveals which source dominates the overall uncertainty at each location which cannot be obtained by Figures 12 and 13 independently. In the uncertainty decomposition maps, areas dominated by data uncertainty suggest that improving the quality or representativeness of the input evidence layers would most effectively reduce uncertainty. Conversely, areas dominated by model uncertainty indicate that improvements in model architecture, training strategy, or parameter estimation may be more beneficial.

7.2 For this specific study case, prospective areas seem to be in the vicinity of existing mines. Is it related to the location of the negative samples? What is learned specifically for this area from the uncertainty and prospectivity maps from previous MPM.

Re: The predicted high-prospectivity areas are clustered around known Au deposits. However, this pattern is not primarily related to the selection of negative samples. Negative samples were selected from non-mineralized areas while avoiding the vicinity of known deposits to reduce the likelihood of including undiscovered mineralization.

In the study area, previous mineral exploration has already discovered some Au deposits. The proposed uncertainty framework provides additional information by quantifying the reliability of these predictions. Specifically, areas exhibiting both high prospectivity and low uncertainty represent the most reliable exploration targets and can therefore be prioritized for exploration. In contrast, high-prospectivity areas with elevated uncertainty should be prioritized for additional geological exploration investment. Furthermore, the uncertainty decomposition indicates that prediction uncertainty is dominated by data uncertainty rather than model uncertainty, suggesting that improving the quality and representativeness of the evidence layers would be more effective than further optimizing the prediction model. Therefore, the uncertainty analysis not only improves the interpretability of the prospectivity map but also provides practical guidance for exploration risk assessment and future data acquisition.