Dear editors,

This is our reply to the review of the article sand-2024-1 entitled “Numerical uncertainty identification, classification and quantification in radioactive waste management” by Vinzenz Brendler and Solveig Pospiech (both Helmholtz-Zentrum Dresden-Rossendorf e.V., Bautzner Landstrasse 400, 01328 Dresden). We have copied all remarks from the three reviewers into this document, together with our responses (all in GREEN). The changes / additions discussed have been done in the manuscript accordingly.

We are thankful for the comments and are confident that the manuscript has now reached a higher quality.

Best regards from Rossendorf,

Vinzenz Brendler and Solveig Pospiech

RC1

The paper provides a summary of some results of the UMAN project. This is in principle fine, but the contribution of the paper should be stated more clearly. In particular, large parts of the text in most sections are just a copy from the UMAN report without marking them as a quote: For example, L293 - L355 (two pages of text), L196-L246, L262ff, L166-185, large parts (in fact most of the text) of the Outlook (section 5) are also just copied from the UMAN report. This has to be changed! Either mark them as direct quotes, or put them in boxes as extracts from the report which you discuss in the manuscript.

The paper directly results from research work done within the EURAD Strategic Study UMAN. The EURAD consortium explicitly encourages participants to transfer their results from the internal ProjectPlace storage into articles in peer-reviewed journals. This transfer actually was one of the Key performance Indicators where the quality of the work was judged. The authors did that successfully already before with two articles resulting from the EURAD Work packages DONUT and Knowledge Management, respectively:

1) Claret F, Prasianakis NI, Baksay A, Lukin D, Pepin G, Ahusborde E, Amaziane B, Bátor G, Becker D, Bednár A, Béreš M, Bérešová S, Böthi Z, Brendler V, Brenner K, Březina J, Chave F, Churakov S, Hokr M, Horák D, Jacques D, Jankovský F, Kazymyrenko C, Koudelka T, Kovács T, Krejčí T, Kruis J, Laloy E, Landa J, Ligurský T, Lipping T, López-Vázquez C, Masson R, Meeussen JCL, Mollaali M, Mon A, Montenegro L, Pisani B, Poonoosamy J, Pospiech SI, Saâdi Z, Samper J, Samper-Pilar A-C, Scaringi G, Sysala S, Yoshioka K, Yuankai Y, Zuna M, Kolditz O. “EURAD state-of-the-art report: development and improvement of numerical methods and tools for modeling coupled processes in the field of nuclear waste disposal”. Frontiers in Nuclear Engineering 3 (2024) 1437714.
2) Knuuti, T., Tatomir, A., Göbel, A., Franzen, C., Abbasova, D., Arnold, T., Brendler, V., Fuzik, K. “Capturing the state-of-knowledge in EURAD Knowledge Management”. EPJ Nuclear Sciences & Technologies 8 (2022), 37

Thus we do not see a necessity to quote these texts separately as the UMAN report is not an official publication.

Moreover, the transfer into regular journal articles is the only safe way to keep the results over longer time periods. One of the authors (VB) took part in several EU framework programs. All the detailed reports at that time stored in the EU CORDIS portal for projects such as NF-PRO or FUNMIG are now not accessible any longer …

However, the last decision is with the editor, namely if the respective text passages from the EURAD UMAN report have to be put in quotation marks, into separate boxes, be-reformulated or kept as they are.

L59: The term "numerical dispersion" has a very specific meaning, which is not the one intended here. Better call it "effects of computational imprecision". In the same paragraph, it is said that such imprecision, e.g. due to convergence criteria of iterative algorithms, are dealt with validation, qualification and verification. Validation and verification address two very distinct aspects in modeling (physical adequacy of the model and mathematical correctness of the implementation). Please reformulate to remain precise.

We switch to "effects of computational imprecision"

The definition of epistemic uncertainty is here put into the context solely of numerical models. In the beginning of the paper, it is said that the study is limited to "numerical uncertainty", where the term numerical is used in the sense of data with numerical values. This is a broader category than numerical (in the sense of computational) modeling. The overall wording in this sense might need a bit of sharpening to avoid confusion. The definition of epistemic uncertainty can be broadened accordingly. One could also discuss that the practical distinction into aleatory and epistemic (with a view on reducibility) can be context-specific (see, e.g., "Special Workshop on Risk Acceptance and Risk Communication Aleatory or epistemic? Does it matter?" by Der Kiureghian and Ditlevsen). In a RWM setting, generic studies are very different from site-specific characterization, for example, or URLs can be explored differently than a proper repository.

The definitions have been negotiated between different actors within EUARD UMAN to harmonize them over the respective deliverables. But we tried to sharpen the definition that would be appropriate to this paper.

Something seems to be wrong in Tab. 2: there's a column of large amount of knowledge and one for high amount of knowledge. Probably the latter should be low amount of knowledge.

Has been corrected

L237: How would such an internal consistency be ensured? What is best practice in this context?

Internal consistency can be ensured in very different ways, a global approach is not existing but strongly depends on the subject. From the areas where the authors are experts that would require to use e.g. identical thermodynamic data sets and reaction schemes for all (geo)chemical compartments, or to use the same electric double layer model for all sorption reactions, or to apply the same activity model for all reactions in highly saline conditions. A sentences had been added accordingly.

The authors discuss fishbone diagrams and show two examples. Explain, how they are linked in a technically (if they are). In the present example, sorption and RN migration are two linked diagrams. How does one move from one to the other? Is this system set up as a dynamic data base or, at the moment, a

collection of examples that just shows that in principle this is a practical way of ordering uncertainty information.

The fishbone diagrams are used as illustrative examples. Of course many subprocesses in the geochemical domain are interlinked, so parameter sets and their associated uncertainty may not only be propagated but can occur in different circumstances. As an example from the figures 2 and 3, the specific mineralogy (as a function of time and space will not only affect the aqueous phase composition / pore water chemistry, but also the diffusion patterns as well as the surface binding site density relevant for sorption. And the latter itself is part of the overall migration model.

The authors have a background in sorption, reactive transport, radiochemistry, etc. and show examples of the BU approach in the form of fishbone diagrams. Can you add a similar example for the TD approach coming from RN assessment? That would add valuable context to the general concepts discussed in the paper.

A fishbone diagram is not suitable for Top-Down approaches as it shows how different very specific submodels and their parameters (with associated uncertainties) are linked together and can form more generic models. We added this information in the text.

We agree that an example for the TD could be interesting to the reader, but given the length of the paper, we decided to provide only one example for BU. The introduction of the section had been slightly adjusted accordingly.

Typos:

L19: "treating with" -> "treating" or "dealing with" -> done

L71: "known knows" -> "known knowns" -> done

L104: pdfs -> probability density functions -> done

L 191: effect -> affect -> done

RC2

According to the abstract, the paper presents major findings from the UMAN deliverable "Uncertainty identification, classification and quantification". As these results have already been published, the authors' own further reflections are of particular importance for the scientific value of the publication. Where results from UMAN are presented and where the authors' own reflections come into play should therefore be more clearly labelled.

See remark to 1st reviewer

The glossary represents a significant added value of the paper. In addition, the glossary should explain what understanding of risk the publication is based on: Is risk only associated with negative effects that need to be assessed or also with positive effects? What role does the probability component play? Against this background table 1 should be reconsidered.

We understand risk as solely negatively connoted, hence only with negative effects. Although the ISO definition of risk (https://www.iso.org/obp/ui/en/#iso:std:iso:31073:ed-1:v1:en) gives a certain space for a more general understanding, we argue that without the explicit mentioning of the ISO definition the

term risk is always understood as the “negative effects of uncertainty on objectives”. Hence, we don’t see the necessity to define risk in the glossary, also to not overload the glossary. What is meant with “probability component”?

The article is explicitly focussed on numerical uncertainties, but - as mentioned in chapter 4 - does not deal with the mathematical handling of uncertainties. It would therefore be appropriate to further clarify the question of whether numerical uncertainties can be clearly distinguished from non-numerical uncertainties and what advantages the focus on numerical uncertainties brings for the publication. The authors declare that they want to concentrate on parameter uncertainties. In the text, however, much space is given to other uncertainties, in particular model uncertainties.

The paper does not state, that the focus on numerical uncertainties brings any advantages. This focus was not a decision by the authors but followed the workload distribution within EURAD UMAN. There, other teams dealt with non-numerical uncertainties, such as work packages addressing “Pluralistic analysis of uncertainty management related to human aspects” or “Identification, analysis and description of preferences of different actors on uncertainty management options”. We have now used the term “parameter uncertainty” only if it is fitting.

Distinguishing numerical from non-numerical uncertainties is primarily based on whether we have a respective number or not. Of course it is clear, that semi-quantitative uncertainty descriptions such as “lower than” or “higher” can transformed into a numerical support system e.g. based on fuzzy set transformations, see e.g. Bandemer & Gottwald, 1995.

We added also a sentence in the introduction to clarify on this in the beginning.

While there is general reference to RWM at the beginning of the publication, later statements concentrate mainly on the (long-term) safety of deep geological repositories. A clearer definition of the scope of the paper would be helpful.

Has been improved accordingly. We replaced RWM where in infact DGR was addressed. As for the scope of the paper, the first sentences of the introduction is starts with “Deep geological repositories (DGR)…” and we assume that this sets clearly the scope of the paper.

The statement in line 64/65 "Uncertainty creates an uncomfortable position for a large part of the public (anxiety)" is not clearly supported by the social science literature. If the topic is not dealt with in more detail in the paper, this sentence should therefore at least be supported by a reference to the literature.

A generic analysis of such phenomena was first published by Smith and Ellsworth in 1985, see new reference. More RWM-specific thoughts are collected in EURAD Deliverable 10.17 – “Synthesis report of WP UMAN outcomes from a civil society point of view”, namely the EURAD Civil Society seminar outcomes as summarized in chapter 2 “Transparency and public participation in uncertainty management”, see also Hooker (2008) or Eckhardt (2021) for applications in RWM.

At the beginning of section 2.1, it is not clear what the cost-benefit ratio refers to. "Relevance can be inferred from priority for further investigation" can be read as a circular argument.

The cost-benefit-ratio refers to the reduction of uncertainties by additional experiments / site characterizations. This has been updated now. “Priority to further investigation” is not expressed correctly, we changed it into “Priority in the RWM roadmap schedule”, i.e. is it something to be considered very soon or can it wait some decades.

Table 2 should be verified. What is the difference between a large and a high amount of knowledge?

This has been corrected

In line 418/419 there is the statement "Finally, further research activities should elaborate possible ways to treat uncertainties resulting from human behaviour numerically, taking into account that their impact on the safety could be significant." Was this statement made against the background of the relevant social science literature? A corresponding reference would be valuable.

This statement should just tell that this paper is not dealing with human behavior. And the authors doubt that many of these uncertainties can be quantified, anyway. The respective EURAD UMAN deliverables D10.8 “Views of the different actors on the identification, characterization and potential significance of uncertainties related to human aspects” and D10.15 “Pluralistic analysis of uncertainty management related to human aspects” does all least not give hints in that direction.

RC3

General comments

In this paper, the results of the EURAD project UMAN are described with a focus on uncertainties in numerical models. The paper effectively highlights the uncertainties and provides options for managing them, such as using a fishbone structure. However, the paper seems to rely too heavily on a EURAD report, which has essentially been adapted with only minor adjustments to fit the format of a paper. For instance, the outlook states that “A thorough survey in other EURAD deliverables issued so far is recommended for indicators that may help to rank the importance of FEPs and their associated uncertainty.” However, I believe much more literature exists on this topic beyond EURAD deliverables. Additionally, the paper notes that “A revision of ranking and order of various uncertainties might also be necessary for the current EURAD Themes definition.” This again focuses solely on EURAD, while there is likely a wealth of other relevant papers on this subject. Therefore, I believe the paper should reference a broader range of other papers and reports. It is also recommended that the authors review the text, as in some instances, they still refer to “this report” instead of “this paper.”

We added more references expanding the focus beyond EURAD, also adapting the text respectively. See e.g. Galson and Kursheed (2006), NUREG (2003), OECD/NEA (2012), Swiler et al. (2021) or Vigfusson et al. (2007). The wording had been changed from “report” to “paper” consistently over the manuscript.

It is also unclear whether the uncertainties discussed in this paper pertain specifically to the safety assessment or safety analysis of a GDR, or if they are related to numerical modeling in general. I suspect it is the former, but the paper does not make this clear. This distinction should be clarified, and an introduction to the subject might be necessary. Additionally, I believe the paper would benefit from including examples with references. For instance, in section 2.1, the authors refer to textbooks that might provide an in-depth discussion, but they do so without citing specific examples. In other sections, they mention certain methods that can be used. Including a reference to a book or other source in the text would be very beneficial for readers who do not have a background in this area. Lastly, I found that there was little on the different stages of a program and how that would influence the uncertainty.

Indeed, the paper’s focus is set on safety analysis of a deep geological repository, this is now made clearer. More guidance is given by the additional references cited just above.

Concerning the different stages of a DGR, there influence on the uncertainty certainly exists. This was one reason why there are different work packages and deliverables associated to some of them within EURAD UMAN. But a discussion of such effects is beyond the expertise of the authors, unfortunately.

Specific comments

Abstract

R9: Safety analysis is mentioned here but later on in the introduction, it is also referred to as safety assessment. Better use one word and maybe explain what it means. -> done, decided for safety assessment because it is more general than safety analysis

A definition of "safety assessment" is added to the glossary.

R14: I am wondering if "for the report" is correct here. In general, I would refer to it as paper. -> done, all instances with "report" exchanged for "paper".

R17: RWM is not defined. -> done

Introduction

Overall the introduction looks good, but it can be shortened by focusing on what is done in the paper rather than describing in length was is not done. This will help to make it clearer what the paper is about and it will reduce the length of the introduction and make it clearer.

We tried to balance this with the request from the other two reviewers to expand the introduction.

R25: Safety assessment of the DGR? -> done

R26: Here, radioactive waste management for the first time in the introduction, so maybe introduce the abbreviation here rather than later on. -> done

R29:Is the safety assessment the same as the safety analysis in this context? -> done, see previous comment

R31: I am a bit wondering how various stages of radioactive waste management influence the safety analysis.

See remark above

R63: Deliverable or paper? -> done

R68: What kind of decisions?

As examples, we now mention decisions about the site selection, layout of the underground facilities, and composition of the technical barrier

2.0 methodology

I like that the authors start with a list of definitions. I do, however, wonder how their definitions are compared to others.

See answer to a similar remark of reviewer #1.

R103: pdfs? -> done

R151: The abbreviation WMO is already defined somewhere else. -> no, we didn't find an earlier place of usage or definition

R117: Both strategy and structure are red underlined. Is that for a reason? -> we can't see red underlining. If that is the case in another viewer, it is not planned and has not reason.

R130: It is mentioned that they are used for a multitude of complex applications in science and society, but could some examples be given?

We tried to give a few examples, see response to the respective "general comments" of this reviewer.

R186: Again report and what is exactly meant by sources being identified and used in this report are listed in the references section.

"Report" had been changed to "paper". "Sources" had been specified.

R205: Sea level changes instead of water transgression? -> ok, somewhat academic: water transgression may occur without sea level change (e.g. by pushing down continents into the mantle through weight of ice cap) and sea level change might not necessarily induce water transgression or regression, if the level of surface remains the same with respect to water level, either through sedimentation, loss of sediments or changes of continent level. The important part here is that water might transgress onto continental plates, and therefore I would prefer the term "transgression" instead of "sea level change"

R244: I don't think guess is good to say here: it is always based on something.

We dropped guesses. References providing more details on expert's judgment are Mumpower and Stewart (1996) or Nuclear Decommissioning Authority (2017).

R251: Design concept of the DGR?

Of the DGR and its major components, namely the mine layout itself.

R276: What are first two columns in table 4 in this paper use for? Do they serve any purpose?

They name the areas where the respective uncertainties are discussed in the overall RWM framework, they list the subthemes and domains defined in the EUARD roadmap. Is now put into the table caption.

3.0 Categorization of uncertainties

R327: The abbreviation FEP is not explained here (or earlier; only later).

Explanation is now added in line 173

R343: EURAD is currently almost finished. So, is there no final version of the EURAD roadmap?

Reference had been updated.

4.0 Categorization of uncertainties

R375: Not the topic of this report? Paper probably.

Has been changed

R376: RW?

Changed to DGR.

5.0 Outlook: What are the gaps in understanding, quantification and processing of uncertainties?

R394: Task 3 deliverables?

This half sentence has been dropped

R410: Why would it be challenging to convince modellers and programmers? I think most of them want to but cannot do it due to time.

It would need a lot of extra efforts, and it has been discussed at least in the geochemical community for more than three decades without much progress. So obviously, there are a few more obstacles than lack of time. But we have softened the sentence.

Added references:

Der Kiureghian, A. and Ditlevsen, O. (2009). Aleatory or epistemic? Does it matter? Structural Safety 31, 105–112.

Eckhardt, A.A. (2021). Sicherheit angesichts von Ungewissheit: Ungewissheiten im Safety Case, Transens-Bericht, Zollikerberg, CH: https://www.transens.de/fileadmin/Transens/documents/Ver%C3%B6ffentlichungen/TRANSENS-Bericht-01_Ungewissheiten-dher.pdf

Galson, D.A. and Khursheed, A. (2006). The Treatment of Uncertainty in Performance Assessment and Safety Case Development: State-Of-The-Art Overview. Milestone (N°:M1.2.1), PAMINA Project

Hooker, P.J. and Greulich-Smith, T. (2008). Report on the PAMINA Stakeholder Workshop: Communicating Safety Issues for a Geological Repository. Deliverable (D-N°: D2.1.B.1), PAMINA Project

Mumpower, J.L. and Stewart, T.R. (1996). Expert Judgement and Expert Disagreement. Thinking and Reasoning 2 (2/3), 191-211.

NUREG (2003). Proceedings of the International Workshop on Uncertainty, Sensitivity, and Parameter Estimation for Multimedia Environmental Modeling. NUREG report CP-0187, Rockville, Maryland, USA

Smith, C.A. and Ellsworth, P.C. (1985). Patterns of cognitive appraisal in emotion. J. Pers. Soc. Psychol. 48, 813-38.

Swiler, L.P., Becker, D.A., Brooks, D., Govaerts, J., Koskinen, L., Plischke, E., Röhlig, K.J., Saveleva, E., Spiessl, S.M., Stein, E. and Svitelman, V. (2021). Sensitivity Analysis Comparisons on Geologic Case Studies: An International Collaboration. Report SAND2021-11053, Sandia National Lab., Albuquerque. DOI: 10.2172/1822591.

Vigfusson, J., Maudoux, J. Raimbault, P., Röhlig, K.J. and Smith, R.E. (2007). European Pilot Study on the Regulatory Review of the Safety Case for Geological Disposal of Radioactive Waste. Case Study: Uncertainties and their Management: https://fank.fgov.be/de/system/files/case-study-european-pilot-group.pdf