



Trust in Models – Open Letter from the Editors

Ingo Kock, Martin Navarro, Jens Eckel, Stephan Hotzel, and Carlo Dietl

Federal Office for the Safety of Nuclear Waste Management, Berlin, 10623, Germany

Correspondence: Ingo Kock (ingo.kock@base.bund.de)

Received: 13 March 2026 – Accepted: 23 March 2026 – Published: 4 May 2026

1 Introduction

In 2021 BASE headed a short workshop called “What do we need to do to trust in models?” at the first edition of the safeND symposium conference (safeND: “Safety of Nuclear Waste Disposal”; Kock et al., 2021). There, we raised a question which, in essence, is twofold: what do we as scientists actually need to trust in our own and each other’s models? How is it possible to convey this trust to the non-scientific public? The first question was asked while keeping in mind that many uncertainties arise when simulating a repository over 1×10^6 years. Moreover, there are difficulties in proving the adequacy of models. The latter question was asked by natural scientists and engineers with an awareness that it cannot be answered comprehensively without the help of other disciplines.

Because of the broad interest people showed in the topic, three consecutive workshops on the topic of “Trust in Models” (TiM) were held between 2022 and 2025. We would like to thank all of the scientists who joined the fruitful discussion. You were great guests, and we were happy to host you. In this letter, we will present the main points of discussion. We present our personal view and want to stress that, of course, there were contradicting views among the participants which we cannot fully cover.

Primarily, we tried to get a handle on the first question: what do natural scientists and engineers need to trust in our models? Many of the points discussed during the workshops were not novel; in fact, many of them were already mentioned (in principle if not in name) in the outcome of the NEA’s “Methods for Safety Assessment of Geological Disposal Facilities for Radioactive Waste – MeSA” initiative (NEA, 2012). Our goal was neither to equal the scope of the MeSA initiative nor to go beyond it. However, in light of the German site selection process, we felt it necessary to discuss and, if possible, determine where we stand today.

2 Complex models

Since computing power has been increasing for decades, models for process and performance analyses have become more complex. It was noticed that the notion of models being “complex” may have a variety of meanings. In any case, it is worthwhile to distinguish between a visible and a hidden complexity. For example, the complexity of a highly coupled model may remain unnoticed due to a simple model geometry. Increasing model complexity just as an end in itself does not necessarily enhance system understanding. Contrarily, adding more uncertain parameters and choosing values for them may only result in fitting the systems’ behaviour to the expected result. Yet, increasing complexity in an informed way may be used to

- find out which processes are or might be important or sensitive
- carry out consistency and plausibility checks by comparing with simple models.

3 Interaction between process modelling, safety analysis, and performance assessment

In nuclear waste management, process modelling has to consider scenarios and strategies for the safety demonstration. It always has to be tested for its direct applicability to safety analyses. Simultaneously, safety analyses have to incorporate process models already in an early stage of their development. In this context, the complexity of a (process) model has to serve the aims of the safety analysis. “More complex” is not always better – both complex and simplistic models (both process and performance assessment models) are necessary and could be organized in a model hierarchy. Therefore, a dialogue between process modellers and scientists working on the safety case is indispensable. The same is true for the relation to experiments (see also the paragraphs below): we

can only model successfully complex models if we compare and benchmark them with experiments and their respective complexity. Furthermore, a proper error analysis and plausibility check is the base for each model whether it be complex or simple. And, finally, scale variance of processes must be considered – this is particularly true for performance assessment (PA) models, whose assumptions are justified by process models which were originally developed for processes at smaller scales than the PA model.

4 Transparency

The fact that transparency automatically builds up trust was not a view that was consensually shared. For scientists, transparency remains important for guaranteeing traceability. However, when communicating with the public, it is questionable whether a transparent presentation of model complexity and model limitations promotes or hinders the development of trust. The answer to this question is of practical relevance: modellers need to decide how to prepare the results for non-scientists.

Seidl et al. (2024) confirm our observation when they state that “there can also be different criteria for credibility. One would be that uncertainties are presented transparently. So that these are then received positively, i.e. in a way that increases trust, there must be basic trust beforehand.”¹

Seidl et al. (2024) argue that “a particularly careful communication strategy must be chosen.”¹ and that “nevertheless, value must be placed on transparency”¹. We would like to note that the basic problem cannot be solved by a single discipline or a few disciplines, and more efforts should be made to address this challenge. However, as discussed in the workshops, part of the problem evolves from the different perspectives that various disciplines – namely natural sciences and engineering – have regarding transparency. Regarding these disciplines, some recommendations can be given:

- Reasons for modelling choices should be documented, at least for the professional public. This helps, among other things, to assess whether the evaluation of safety based on model calculations shows a subjective bias.
- Code developers often do not document their codes down to every detail. However, they should ensure complete documentation and also record code changes. This should include the applied workflow and auxiliary codes on which the modelling results depend.
- Documentation of the modelling software and the applied workflow (step by step) and also a proper linkage to the published literature are key for transparency. Only then can gaps within the recent research be identified and the quality of one’s own research be shown and evaluated.

¹Translated from German by the authors

- Models (and their results) depend on the model assumptions. The physical or numerical meaning of model assumptions is not always explained in detail. For example, it may not be obvious that the scaling of model elements makes the implicit statement that processes on sub-scales can be homogenized. Because such implicit assumptions need to be checked, they should be identified and documented if possible.
- Analogously to numerical modelling, there should be a comprehensive description of experimental setups, procedures, and conditions.
- The FAIR (Findable, Accessible, Interoperable, and Re-usable) principle, as described for research data (Wilkinson et al., 2016), should be applied to numerical models (e.g. input–output data sets) where possible. In addition, if the software is open source, the framework would correspond to the principle of transparency in the German site selection legislation (StandAG, 2017).
- Data sets and parameters should be versioned to be similar to the versioning of source codes.
- A proper linkage to the published literature is a key element for transparency. Only then can gaps within the recent research be identified and the quality of one’s own research be shown and evaluated.

5 Benchmarks

In the discussion, it was generally acknowledged that code benchmarks are necessary for knowledge build-up. However, code benchmarks can be improved:

- Deviations within benchmarks are often caused by ill-defined benchmark problems. Very basic assumptions (e.g. the compressibility of water and other fluid parameters) are often not part of problem descriptions. This may lead to strong deviations in terms of modelling results. Well-defined problems could help to avoid large initial deviations and avoid misperceptions of model quality. Well-defined problems should include the definition of constitutive laws and the obligation to use them. In case one wants to test different constitutive laws, this should be done with different sets of models and should be clearly communicated.
- Benchmark results can deviate if modelling groups use different spatial and time discretizations. The dependence of discretization within benchmarks could be better explored within benchmark projects.
- With respect to transparency, the scientific community could be involved in the benchmarking process: data and results need to be provided openly.

6 Experiments

Choosing physical models and model parameters for numerical models also depends on experimental findings. Therefore, the trustworthiness of numerical models is also determined by the trustworthiness of the experiments. It is therefore important that safety analysts understand the experimental results they are using and the associated uncertainties. A discussion with the experimental scientists can aid in this understanding (see also “Interaction between process modelling, safety analysis, and performance assessment” above).

- Analogously to numerical models, experiments could be successively expanded from simple to more complex ones.
- Experiments help in identifying which processes might be possible in the repository system, although experimental phenomena need not necessarily be relevant to the system.
- More attention should be paid to the interaction between experiments and safety analyses:
 - Experimental results could contribute to the construction of performance assessment models by providing information on adequate model simplifications, as well as the uncertainties and sensitivities of the experimental system.
 - Performance assessments should identify which modelling assumptions require more experimental support and formulate experimental demands as precisely as possible.
- Experiments are useful for validating models and for the calibration of models; they can close calibration gaps in models.
- Experiments – as models – need a quality control: they should be in correspondence with the state-of-the-art theory. It is best if they are based on an unequivocal statistic and a robustness which can be founded in round-robin tests and other benchmark measures.
- It should not be forgotten that experiments take place on different timescales – very often as a result of their nature as laboratory or in situ experiments. We should always have the period of consideration for nuclear waste disposal, i.e. 10^5 or 10^6 years, in mind when doing experiments and models.

7 Human reliability

Classical methods of quality control help in reducing modelling errors owing to human reliability.

- Verification, standardization, and validation of workflows (including data management, pre-processing, and post-processing) should be given similar importance to the verification and validation of simulation codes. The application of artificial intelligence (AI) (see below) can be helpful here as well.
- The invisible routines of the workflow used for data conversion, extraction, evaluation, and presentation should be documented and subject to quality assurance. The reason for this is that modelling results can depend heavily on these routines. Moreover, good documentation is necessary for the reproduction of modelling results.
- Data management measures should be established. As large data sets evolve and are subject to revisions, they may become erroneous and inconsistent. Data management should therefore include information on data origin, data dependencies, and decision history.
- Code coverage has to be analysed. Parts of code that have not yet been tested cannot be considered to be quality-assured. Code coverage analyses help to point out which parts of the code remain to be tested. However, the fact that a part of a code has been executed does not prove that the code is correct because it has not been executed with all possible variable states of the code.
- An important measure of quality control is cross-checking modelling results by repeating the simulation with complementary models or with a different modelling group (these can be organized intra- or inter-institutionally). This is a way of implementing a four-eyes principle. It is recommended that implementers, regulators, and their respective consultants do not use identical codes – in case they use only one modelling software. In case they can apply more than one code, it can be useful if the regulator also works with the implementer’s code, simply to understand its software.
- It is necessary to preserve the knowledge of code usage and development at the scientific institutions by means of a good knowledge transfer from one generation to another.

Due to the complexity of models, human reliability will always be an important issue in building confidence in models. Human misinterpretation of modelling results is a related aspect. This includes a misjudgement of the success of performance assessment models, which, in fact, cannot be inferred from the model result but only from a close scrutiny of the model itself, the data it uses, and the given uncertainties. A model may gain an inappropriate reputation if the fact that the model produces some result is mistaken for the fact that the model produces the correct result.

8 Artificial intelligence (AI)

The importance of AI will increase in relation to the modelling of repository processes. Implementers, as well as regulators, need to independently assess impacts and future research needs. Without being comprehensive, discussions are going in the following directions (all bullets relate to modelling for repository systems):

- AI-based surrogate models will likely be state of the art in the future and, at least, complement physics-driven models. Specific research and development are necessary to establish requirements for machine-learning (ML) and AI usage.
- A combination of data-driven AI and physics-driven modelling is important. The data used to train the AI may be the result of underlying physical processes and/or physics-driven modelling.
- The importance of training and classifying data sets increases and will play a role in modelling since every model's quality is limited by the underlying data sets.
- AI will probably never replace human-made models, but it will, for sure, complement them and help to accelerate modelling processes.
- Language tools can be a great help for the syntax check of codes. ML can assist in identifying data gaps.
- Maybe, in the future, AI can be used for uncertainty quantification and to avoid error propagation in large-scale (PA) models.
- Finally, the application of AI can lower the hurdle for unexperienced modellers in developing their own modelling ideas and in getting familiar with coding.

9 Conclusions

The fact that scientists doubt models does not mean that models are not trustworthy. A critical attitude is essential for the improvement of models and modelling results.

If models have been successful for short-term processes, it does not necessarily mean that they will provide correct extrapolations for long-term processes. However, even under these circumstances, model confirmation increases with every new failed attempt at proving the inadequacy or incorrectness of the model and the input data. Consequently, the confirmation of modelling results is not only a feature of models but of the entire modelling process which, for this reason, should be documented. In fact, whether a model appears to be trustworthy or not does not become apparent from the presentation of modelling results alone. It also depends on how much effort has been spent to arrive at a certain modelling result. This effort should be communicated too.

Building trust in models and modelling results remains a challenging field of research from both a practical and a theoretical point of view.

Data availability. No data sets were used in this article.

Author contributions. All of the authors contributed equally to the original draft preparation and to review and editing. MN contributed to the conceptualization of the original workshop idea. IK contributed to the supervision of the activity.

Competing interests. At least one of the (co-)authors is a member of the editorial board of *Safety of Nuclear Waste Disposal*. The peer-review process was guided by an independent editor, and the authors also have no other competing interests to declare.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. The authors bear the ultimate responsibility for providing appropriate place names. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Special issue statement. This article is part of the special issue "Trust in Models". It is a result of the "Trust in Models" workshop series, Berlin, 2022–2025, Germany.

Acknowledgements. Our big "thank you" goes to all of the scientists and institutions who have contributed to the past TiM workshops with their presentations, opinions, and arguments.

Review statement. This paper was edited by Sarah Glück.

References

- Kock, I., Navarro, M., Eckel, J., Rücker, C., and Hotzel, S.: What do we need to trust in models?, *Saf. Nucl. Waste Disposal*, 1, 303–302, 2021.
- NEA: Methods for Safety Assessment of Geological Disposal Facilities for Radioactive Waste: Outcomes of the NEA MeSA Initiative, Radioactive Waste Management, 6923, OECD Publishing, Issy-les-Moulineaux, France, 238 pp., ISBN 978-92-64-99190-3, 2012.
- Seidl, R., Becker, D.-A., Drögemüller, C., and Wolf, J.: Kommunikation und Wahrnehmung wissenschaftlicher Ungewissheiten, in: Entscheidungen in die weite Zukunft: Ungewissheiten bei der Entsorgung hochradioaktiver Abfälle, edited by: Eckhardt, A., Becker, F., Mintzloff, V., Scheer, D., and

Seidl, R., Springer Fachmedien Wiesbaden, Wiesbaden, 313–336, https://doi.org/10.1007/978-3-658-42698-9_15, 2024.

StandAG: Act on the Search for and Selection of a Site for a Disposal Facility for High-Level Radioactive Waste: Site Selection Act – StandAG, BGBl. I, 26, 1074–1100, 2017.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., Hoen, P. A. C. 't, Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B.: The FAIR Guiding Principles for scientific data management and stewardship, *Sci. Data*, 3, 160018, <https://doi.org/10.1038/sdata.2016.18>, 2016.