RC1

This is a nice summary article of aspects of the extensive and structured work done in the KOMPASS and MEASURES projects that is, from my point of view, worth publishing. The research done in these projects is highly relevant to our field. Even though specific results are not discussed in detail as the article rather gives a synopsis and discussion/outlook of the projects themselves, I have some minor comments that could further improve the article. These are some small wording aspects on the one hand and some requests for discussing those results a little more that are selected for presentation in the article.

We are sincerely thankful for the review and the comments that helped to improve our paper. Below, we outline how we address the comments in the manuscript.

- In the glossary, I would like to see an aspect added to the definition of the definition of "constitutive model". The description of a specific physical system is usually not achieved by constitutive models alone, but still includes the (more general) balance equations. I'd therefore suggest to rephrase slightly while maintaining the authors' direction: "Conceptual/mathematical description used as a closure relationship for a specific physical system".

We address this comment together with the first comment of RC2, that goes in the same direction.

We changed the definition to: “Conceptual/mathematical description of the physical behavior of a particular material or substance.”

- Figure 2: I know this is not the authors intention, but one could interpret the linear triangles on the bottom such that uncertainties will be reduced to zero and we'll end up with a fully robust and accurate model if only we have enough follow-up work. There are diminishing returns here, and residual uncertainty cannot be avoided in geotechnics for a number of reasons. This could easily be indicated in the figure (curved triangles or otherwise).

Thanks for the clarification. You’re completely right and we do not intend to say that all uncertainties will be eliminated, only that they will be reduced to a residual. We adapted the figure.

- L123/4: Can you say something regarding the in-situ applications, ongoing or planned? Are there works that can be referenced?

We added the information that there is a KOMPASS backfill body installed in the Sondershausen mine that is monitored for in-situ data.

- L141 is somewhat related: How can the improvement of the constitutive models be demonstrated on the virtual demonstrator, if there's no real-world equivalent at the moment? Say a few words on how this goes beyond just a demonstration of observed differences between different constitutive models and what arguments are used to show the improvement.

Thanks for the comment. We added a few words for clarification:

The continuous improvement of the constitutive models is visualized using a virtual demonstrator representing a generic backfilled drift in rock salt (Figure 4a). The virtual demonstrator is envisaged as a tool to visualize and quantify calibration progress, not to compare simulation results with in-situ data. After the investigation of each new factor or process is completed, the virtual demonstrator is applied, and the evolution of porosity – identified as the most important process variable – is evaluated by comparing the results obtained by different teams. The use of the virtual demonstrator ensures that all models reproduce the simplified behavior of rock salt in a consistent manner, such that the only differences between the models arise from the constitutive description of crushed salt. In this way, changes in individual team results and the bandwidth of the predictions associated with the various crushed salt models can be systematically assessed.

- L135: "uncertainties are avoided by using different formulation options [...] and using different methods". Clarify how that avoids uncertainties. It can help capture model-/method-related uncertainties, thus giving you a clearer picture of the scatter when with only one model you'd be flying blind. Is that what is meant here or is it something else?

Thank you for the questions. You're right with your statement, that's what is meant here. We changed the word "avoided" to "captured".

- L148: Lab data is not used to verify the models according to the definition used in the glossary (which presents a definition to which I agree). Perhaps clarify or take out.

We removed the word "verify".

- L149: "benchmarking supports consistency checks between experiments and simulations". In what sense? You already use the terms confirmation and validation for comparison of experiments and simulations. If you use a specific meaning of benchmarking here, you might want to add it to the glossary. From section 4 it seems to me you use it here as a means of getting a handle on model uncertainty, but I'm not sure.

We changed the sentence to: "benchmarking supports consistency checks between different constitutive models".

- Fig. 5: Very nice figure, particularly the bottom part. I guess your sigma_v is deviatoric stress? In that case use sigma_d as in text (appears 5 times in the figure, if I didn't miss anything).

Thanks for pointing it out. We change the text accordingly.

- Fig. 6: Volumetric behaviour seems to scatter less between models than deviatoric behaviour (here: axial), which might be interesting to comment on if you have an insight into this. Are these all best-fit results for the models?

Thanks, we added a paragraph with more information:

Overall, the model results represent the best fits achievable with the currently available ("to-state") model formulations. For volumetric strain, the simulations show good agreement across modeling teams and with experimental data, indicating a consistent representation of bulk compaction processes. In contrast, a larger scatter is observed in the axial strain predictions. This variability can be attributed to differences in calibration priorities – where volumetric strain was generally emphasized over axial strain – as well as to known deficiencies in some constitutive formulations, such as an over-sensitivity to changes in deviatoric loading. Additionally, variations in the representation of shear-induced creep mechanisms and stress redistribution contribute to the observed discrepancies. These results underline the need for further refinement of model formulations and calibration strategies to achieve a more balanced and robust reproduction of both volumetric and deviatoric creep behavior.

- Fig. 7: The creep model for the rock salt is identical for all groups, allowing the comparison of the effect of the crushed salt model. Volumetric behaviour was well reproduced in Fig. 6. Here, quite some scatter remains. In the text, it is mentioned, that calibration reduced the scatter. This is true only in a porosity range below 10%, while in the rest of the porosity domain scatter even increased, with one model standing out. The tests covered 3 to ~17%. The maximum spread between the models occurs in the right graph at roughly those 17% with an order of magnitude spread in stress. Here, the discussion on page 11 should be a bit more detailed and nuanced, rather than simply speaking of improvements, as this is part of the motivation for your follow-up work.

Thanks, we agree that more information helps the reader to follow our statements. We added the following paragraph:

The demonstrator aided in identifying areas of focus for necessary calibration and model development. In the state of KOMPASS-I (left panel) the largest deviation in results is observed around 10 % porosity, determining the focus on the mid-to-low porosity level. In KOMPASS-II, the TUC-V2 triaxial compaction test covered a porosity range from 16 – 3 % porosity, and by incorporating this data the spread in results is reduced for the mid-to-low porosity area (right panel). However, the results highlight the necessity to

focus on high-porosity/low-stress regime, which is currently addressed within the MEASURES project.

- The reference list consists mainly of reports. List some journal papers from peer-reviewed publication on the topic of crushed salt behavior as well.

We added more references that are peer-reviewed.

- Personally, I'd refrain from using informal formulations like "family" (better: team or project members) and "the topic is huge" (better: broad, extensive). There are good reasons for scientific articles being formulated on the sober side, even if it can feel boring.

Thanks. We changed the heading of Table 2 into “MEASURES project members that are not listed as authors” and “huge” into “broad”.