We would like to thank both referees for the thorough review and valuable comments which helped us to improve the quality of the manuscript.
We answered your remarks below one by one and made the corresponding changes in the revised version of the manuscript. We have used the following colors to mark changes suggested by the individual referees:

- Referee 1: blue color
- Referee 2: orange color

Please note that the numbering of sections, figures, tables, equations, and lines may have changed in the revised version of the manuscript. Whenever we refer to one of these objects in our response, we refer to the original (non-revised) manuscript if not explicitly stated otherwise.
It was suggested after submission to ensure that readers with color vision deficiencies can correctly interpret our findings. Therefore, we have also adapted the color scheme of our plots to the "colorblind" map (`https://gist.github.com/mwaskom/b35f6ebc2d4b340b4f64a4e28e778486`). We hope that this change improved upon this fact. However, it is not trivial to find a perfectly suitable color scheme that can be used to represent categorical data with a larger number of categories as present here.

## Response to Referee 1

Note: The comments were extracted from the provided annotated pdf for the purpose of replying to the comments.

- p. 1

  Highlighted: El-Husseiny, 2020

  Comment: this paper is about porosity (and cohesive effect), not permeability. The one about permeability is El-Husseiny, 2021: El-Husseiny, A. Unified Packing Model for Improved Prediction of Porosity and Hydraulic Conductivity of Binary Mixed Soils. Water 2021, 13, 455. https://doi.org/10.3390/w13040455

  Thank you for noticing, this has been corrected.

- p. 17

  Highlighted: size range of 0.02 mm - 200 mm.

  Comment: it might be nice to show few examples of continuous particle size distribution if available

  We added a new figure (in the update manuscript: Fig. 10) to display ten of these particle size distributions.

- p. 17

  Highlighted: The initial porosity $n_{0,i}$ was taken as 0.37 for all size classes, corresponding to the average of the laboratory measurements with uniform packings with similar sediment (Frings et al., 2011).

  Comment: was that based on the data in Figure 9? the average porosity seems lower for sediments with grain size larger than 0.15 mm. If you have measurement of porosity for uniform sediments, please use and plot on top of data in figure 9.

  The value of 0.37 was based on the average of all porosity measurements shown in Fig. 9, and is thus in line with the value stated by Frings et al., 2011 and used in Sec. 5.1 for the parameterization of the cohesion model. We added a statement to make this choice clearer. We agree that a more refined choice of $n_{0,i}$ based on uniform packing measurements for the different size classes could improve the outcome. Here, however, we decided to go with the single value for simplicity reasons.

- p. 17

  Highlighted: $d_c^{coh} = 0.5$mm

  Comment: should be 0.15 mm

  In fact, for the data obtained from laboratory measurements by Frings et al. (2011) and presented in Fig. 9, we found $d_c^{coh} = 0.5$ mm. We added this value to the description of Fig. 9 for clarity. Differences to the value of 0.15 mm found in literature, and also reported in Sec. 5.1, might be due to different material or the wet packing conditions. A better insight into the effect of cohesion would be desirable but would require further experimental studies.

- p. 18

  Highlighted: They are significantly smaller than the Rhine

  Comment: significantly smaller in terms of what?

  We added the information that this was meant to be in terms of width, depth, and discharge of the rivers.

- p. 19

  Highlighted: same region

  Comment: you mean same range?

  Yes, we meant "range" and clarified this in the text.

- p. 19

  Highlighted: hint towards packing conditions that result in a rather loose packing state.

  Comment: is there any support to this (Bes and Galabre River) having loose packing or more contribution from finer grains compared to Rhine river?

  We are not aware of any quantitative analysis of the packing condition of the rivers. Note, however, that in the present study, we compared the model predictions to laboratory measurements and not to field measurements. There, the preparation of the sample is known to have a significant influence on the packing condition, e.g., how the sediments have been added to the container and if the container had then been shaken or not. Unfortunately, no information regarding this process were given, neither for the Rhine data set nor for the other two data sets.

  We added a statement to the text to clarify this.

  Regarding the fine grain content, we would argue that systematic differences that might be present between the data sets are taken care of by the packing model and the cohesion model. They should not result in a significant reduction of the prediction quality.

- p. 20

  Highlighted: available

  Comment: add "." at end of sentence

  Done.

## Response to Referee 2

### Clarification needed

- Introduction: Improve the explanation and illustration of interaction between size classes. The nature of the interaction was not clear to me after reading the manuscript, and the recommendation below is based on what I came up with after shortly thinking about the problem:

- 58 State some of the interaction types.

  We added more information about the modelling idea of the interactions, the interaction types and their effect to the text.

- Figure 1 Reduce the size ratio of subplot b, where interaction takes place, to about 1:3, currently, the ratios on all subplot is very large (1:20), which does illustrate the filling and occupation well, but not the interaction.

  We decided to keep the size ratio to not imply that the interaction would only happen for medium size ratios, i.e., to keep it consistent with the two other subplots. But we agree that without further explanation the interaction is not clearly illustrated. We therefore added additional information to the text that refers to the figure in order to explain how the interaction is visible in the figure. We also visually marked the effects in Fig. 1.

- Figure 1 Add a subplot of the porosity of a binary sphere packing vs. size ratio for a couple of volume fractions. The curve has likely a maximum of around 0.36 at $d_1/d_2 = 1$ and a minimum close to $0.36^2$ = 0.13 in the limit $d_1/d_2 \to 0$ for appropriate volume fractions.

  We agree that the relation of size ratio and porosity for a binary sphere packing is generally interesting to visualize and reason about. For the target readership of the present paper, i.e., people interested in fluvial sediment, this might be less relevant since the size ratios that can be found in rivers cannot be influenced and only exhibit a relatively small spatial variability. On the other hand, the volume fractions of the sizes class can vary significantly. This is one of the reasons why the volume fraction is typically the main quantity that enters packing models, followed by the size ratio whose effect is captured by the interaction functions. The illustrations in Fig. 1 also target the effect of the volume fraction instead of the size ratio, so a plot that shows another effect might not be appropriate there. Finally, such a porosity vs. size ratio plot would need to use one of the packing models to be drawn and might thus not be accurate. For these reasons, we decided to not follow this idea despite its interesting point of view. Note, however, that this information is already visible in Fig. 3 if one picks a specific volume fraction and compares the lines representing different size ratios.

- Figure 4 It would be insightful to retransform the functions into a form that their physical meaning can be interpreted (similar to the curve recommended above) and which facilitates a comparison between the models. In the given form, the physical meaning of the interactions is unclear, and the apparent inverse definition of the functions of the CPM and LMPM complicates their comparison.

  We fully agree that a mathematical representation that enables a clear physical interpretation and direct comparison of the models would be very helpful and desirable.

  In fact, this was tried during the initial preparation of the manuscript. We found that a physical meaning cannot in general and for all models be intuitively extracted nor visualized by the interaction functions alone. The reason is that in many cases the models are derived in a rather ad-hoc fashion following arguments for certain limiting (filling and occupation) and intermediate cases. Furthermore, the transformation of a model from one formulation to another (e.g., from specific volume to packing density) is often non-linear (see Eqs. 1-4) which results in lengthy expression for the transformed model and introduces an ambiguity as to which terms should be taken as the transformed interaction functions that one would like to visualize.

  Theses realization resulted in our statement in line 195 that "those interaction functions should usually be seen in the context of the model itself". We thus decided to report the models as they were originally derived to avoid introducing the aforementioned ambiguities for the sake of such a comparison.

  We note that such an in-depth and more theoretical comparison would be beneficial in order to extract common features and differences between the models. This knowledge would be helpful in the future to come up with improved models. Here, we focused on the application-based comparison.

- Introduction: It's worth mentioning that the packing of sediment has relevance in the environmental sciences far beyond fluvial sediment, for example, it also determines the pore size distribution and with it the hydraulic conductivity, and because it is linked to the problem of compaction and hence land subsidence.

  Thank you for that additional piece of information. We added a corresponding statement to the beginning of the introduction.

- Introduction: It's worth a disclaimer that the study implicitly assumes that the packing density for a given size and shape distribution of the sediment is unique. This is not the case, as sediment can compact.

  It is true that it is a limitation of most of the investigated packing models that the packing state is not directly taken into account. As we illustrated in Fig. 11, it may be provided via the initial porosity values that also differ for loose and compacted packings (see also the discussion later on in this response about our notion of the initial porosity). In the CPM, the compaction index $K$ would be another option to introduce knowledge about the packing state. We added this point to the discussion in Sec 5.4, i.e., the new section 6.

- 105 Define $d$. It is probably the sieve diameter, it is only introduced as "size" which is ambiguous. Also mention how $d$ is determined for virtual grains in the computer simulations, as their size is probably not determined by sieving.

The definition of $d_i$ is given in Eq. 5, and is the geometric mean of the sieve size interval as stated in the text. The values for the investigations in Sec. 3 are then given in Tab. 1. We added a statement to Sec. 3.1 how the size of a virtual grain is determined.

- 112 "initial porosity" as a parameter is confusing, as the initial porosity should not matter. What the authors probably mean is the porosity for packing sediment of a homogeneous size, i.e. the packing with one size class only. If this is the case, then change "initial porosity" to "homogeneous packing porosity".

  Thank you for making us aware that the terminology might be confusing. The term "initial" is actually taken from the original papers where the packing models have been derived (Yu & Standish (1991)) and we adopted this notation in order to be consistent. We still think it is advisable to follow that terminology but added a statement to the introduction and to Sec 2.1 to clarify that a "homogeneous" packing is meant here.

- 226 Discuss why the threshold of 150 um for sediment to become cohesive is considerably larger than the usual threshold at the sand-silt transition at 65 um. Many sand bed rivers have a sizeable portion of bed material in this range without that cohesion is considered relevant.

  The threshold for which cohesion presumably starts to play a noticeable role for porosity was found to be 150 um by Zou et al. (2011) and El-Husseiny (2020), whereas based on the laboratory measurements from Frings et al. (2011), shown in Fig. 9, we determined it to be at around 500 um. We agree that an in-depth investigation of this difference, and also the difference to the value of 65 um that you mentioned, would be very much desirable. Ideally, a better understanding of this effect would remove uncertainties regarding the choice of $d_c^{coh}$ and render the packing models more robust and universally applicable. We would assume that different materials or the wet and distinct packing conditions in the experiment could result in such deviations. Due to the limited data that we have available, especially with respect to the effect of cohesion on porosity, we are unable to provide a profound discussion and consequently decided to refrain from doing so.

- 216 The same shape is assumed for sand and gravel. In my experience, gravel deviates much more from spheres than sand, while sand tends to have sharp edges. Should the interaction function thus not only depend on the ratio, but also on the (geometric) mean of two interacting size fractions? Could this be one reason that the parameters of the interaction functions are site specific?

  The interaction functions definitely depend on the grain shape (see e.g. Yu et al, 1996) which also explains their different parameterizations available in literature (see e.g. Liu et al., 2019). Therefore, if the shape in a specific site is significantly different from the shapes that we investigated, the parameterization will most likely require an adaption. Furthermore, it is also correct that the assumption of having the same shape for all size classes might not hold in general. For the relatively narrow size range of 2.8 - 31.5 mm that we could investigate with our simulations and for which we had CT scans available, the preliminary studies reported by Rettinger et al., 2022 indeed showed that shape does not depend on the grain size. This formed the basis for our assumption. However, we fully agree that this might not be applicable for smaller grains for which we have no data available to check that. Consequently, to improve the model, one could introduce the actual size of the two interacting size classes as an additional dependency of the interaction functions, in addition to only the size ratio. This might be an interesting path for follow up work and we therefore added this thought to the discussion in Sec. 5.4 / now Sec. 6. Potential difficulties might arise in the drastically increased amount of required experiments to be able to determine the parameters of these extended interaction functions.

- 218 Which resolution had the CT-scans? Did they resolve the fine particles?

  The resolution of the CT scan was 0.25 mm. This finest grains under investigation had sizes of 2.8 mm for their smallest axis. While this resolution of at least 12 pixels along each axis does probably not permit an exact representation of all surface features, the overall form would still be captured. This applies even more so for the larger grains.

- 220 State the distribution, standard deviation and covariances between the shape parameters, can be done in the appendix or supplement.

  This information together with other details about the form can be found in our previous paper (Rettinger et al., 2022) that is available as an open access article. For completeness, we added the information about the standard deviation of flatness and elongation and their weak correlation to the text in Sec. 3.1.

- 381 The packing state (loose vs. dense) of the initial porosity is an input to the "initial porosity". This makes no sense as the sediment is repacked. What is probably meant is the packing state of the final packing. Consequently, I would also refer to it as the "final packing state" and remove the arrow to the "initial porosity". The final packing seems to be related to the compaction and hence non-uniqueness of the porosity, as I mentioned above and therefore worth to be discussed.

  The packing state itself is admittedly a rather ambiguous input that is difficult to quantify, especially on site. Still, it is an important property of any packing and, as you correctly noted, introduces variations in the obtained porosity. In this work, and for consistency among the different considered packing models, we included the packing state into the initial porosity. This means that the initial porosity values that are supplied to the models should reflect the packing state in which the packing will be. Only in this way can we ensure that the packing models yield the same porosity as the initial porosity for the special case of a having only a single size fraction. We added a statement to Sec. 2.1 where $n_{0,i}$ is introduced to clarify this important point, and picked it up in the discussion in Sec 5.4 / now Sec. 6. Note, that this is different from the notion of the "initial porosity" that e.g. De Larrard (1999) used for the CPM, who assumed it to be the rather loose homogeneous porosity and introduced the compaction index $K$. This notion might be more in line with your argument but the question remains how this loose initial porosity can be determined. To transfer it to our here applied notion, the modification given by Eq. 15 is necessary.

- 401 Why does the simulated packing density of homogeneous sediment differ between the simulation and the lab? Is this due to variation within the size class or does this indicate a systematic error in the models?

  In our previous study (Rettinger et al. (2022)), we thoroughly compared the simulations to laboratory measurements from Liang et al. (2015) to validate the simulation approach. We found a very good agreement using the same CT scans as here. Due to the relatively small size of those experiments and detailed information about the packing process, we could reproduce the experimental conditions very well in the simulations.

  In the present study, we found differences in the determined porosity values for homogeneous packings between the simulations and the laboratory experiments by Frings et al. (2011). Those experiments were carried out in significantly larger containers, resulting in much more particles, that prevented a 1:1 reproduction within the simulations. Instead, we used periodic simulation setups that removed the porosity-increasing effect of the horizontal walls from the simulation. Those are however present in the experiments which might thus exhibit larger porosity values.

  Additionally, the packing conditions of the experiment were not given. In the simulation, we reused the same conditions as for our previous study which we would describe as relatively dense due to the applied shaking. This is another factor that might result in different, i.e., lower, porosity values in the simulations. Other factors like a limited number of samples of the real sediment might also introduce variations. All in all, we would not expect systematic errors in our simulations. We added more information about the origin of this mismatch between simulations and experiments to the text in Sec. 5.2.

- Table 2 It's unclear how the model parameters can be derived from the four inputs stated in Figure 11. Some information on the fitting, i.e. what kind of calibration data is required and how it can be best obtained would be helpful for readers who want to apply the models. This could be provided in a supplement. It seems the parameters are fitted to measured porosities, but this makes predicting the porosity a chicken- egg problem. If the porosity has to be extensively measured at each field side, one could directly predict the porosity based on quantiles of the grain size distribution with a non-linear model or neural-network. It would also be interesting to use such a simple fit as a Null-model or benchmark for the packing models.

  The process how the model parameters in Tab. 2, i.e., the coefficients of the interaction functions, are derived is carried out and explained in Sec. 3. This also demonstrates the data that is required to carry out this procedure. We rechecked these explanations and adapted the text to improve the understanding and to add details to the procedure. The determination indeed requires a fitting to measurement (or simulation) data. However, these measurements must be carried out for bidisperse, and thus artificially generated, packings in the laboratory instead of on-site. Laboratory measurements exhibit significantly less effort than on-site measurements, and can even be replaced by simulations as shown in the present paper. We consider this a particular advantage of the packing models over other models that directly try to fit or learn data from on-site measurements. In fact, Frings et al. (2011) compared the LMPM with a parameterization taken from literature, i.e., without prior calibration, to other models that rely on statistical quantities of the size distributions. This comparison included predictors that were available in literature and one that was specifically fitted to the available data. Even in that case, the LMPM was found to be superior, demonstrating the capabilities of these packing models.

**Textual recommendations**

- I suggest restructuring the manuscript slightly:
  - Move the cohesion model (section 5.1) forward to the end of section 3 (methods section). Currently, it's awkwardly embedded in the result section.
  - Give a brief and clear overview of the models in the beginning of section (3), similar to lines 447-450 in the conclusions. Move the first part of the discussion (lines 390-394) including Figure 11 at the start of section 3, as it details the general overview of the model structure and parameters. This will be helpful for readers who are no experts on sediment packing models.

  We took your suggestion as a motivation to rethink the manuscript's structure in terms of logical sequence. As a result, we made the following changes, mostly in line with your suggestions: We moved Fig. 11 to Sec. 2 where we introduced a new subsection (now 2.2) to provide a general overview of the packing models, their input and output and general flow of information. This should provide readers with a gentler introduction to the topic. Furthermore, since the discussion in Sec. 5.4 has grown due to the reviewer comments, we moved it to its own section, now Sec. 6. We also changed the title of Secs. 4 and 5 slightly to better reflect the structure and content.

  We decided to keep the description of the cohesion model in section 5.1 since it uses data from the same experimental campaign also mentioned in Sec 5.2 to show the effect of cohesion and to parameterize the model while being introduced. Since it is an actual extension, only to be used for cases where cohesion becomes important, we feel that it is justified to keep it separated from the rest of the model description, given in Sec. 2.

- Use active voice, i.e. "we ..."

  We appreciate that there seem to be different opinions about using the passive voice or, as suggested here, the active voice in scientific texts. In fact, in the review process of a different article, a reviewer suggested the exact opposite, i.e., to avoid using the active voice.

  We checked the journal's submission guidelines and guideline books like "Handbook of Writing for the Mathematical Sciences" by Higham, and could not find a rule that either is generally considered a problematic writing style.

  We therefore decided to keep the current style.

- Use present tense where possible to make the results a refreshing read. The persistent use of past tense gives the impression the results were obsolete or outdated.

  We again appreciate that there are different perceptions about the adequate use of the present or past tense when presenting the results, as this had also been a vivid discussion among the authors while preparing this manuscript. Ultimately, we decided to use the past tense for activities that already happened in the past, i.e., most of the experiments, simulations, evaluations, etc., and the present tense only for statements that hold in general. This distinction is not always unambiguous so we checked the manuscript again and moves some statements to the present tense.

- For consistency, use "grains" instead of "particles" throughout the entire text. Currently the terms are jumbled without following a conceivable logic.

  Done.

- Use "evaluate/evaluation" instead of "validate/validation" throughout the text.

  In the context of this paper, we used the terms "evaluate/evaluation" to denote the generic action of obtaining results, e.g., to determine porosity via a porosity predictor, like the packing models or a simulation. On the other hand, the terms "validate/validation", and also "verification", imply a comparison to other data in order to draw conclusions about the accuracy of the model. We would argue that this difference justifies using these two distinct terms in the text for improved clarity.

**Minor textual**

- Figure 4 A logarithmic scale for ratios is more appropriate.

  Done.

- Figure 6 prediction models → model predictions (lines) ... Rhine sediment (dots)

  Done.

- Figures 6 The figure would be more easy to read when the plots were labelled MPM, CPM, LPM, similar to figure 8.

  We made the labeling of all figures consistent and always added the abbreviation of the packing model.

- 3,472 "theoretical" → "algebraic". Theoretical implies the models would be derived from first principles, yet they seem to comprise of heuristics and fitting curves. Algebraic makes sense, as they provide a direct prediction without simulations or solving a differential equation.

  We agree that "theoretical" might be misleading for the reason mentioned by you. However, we wanted to set these packing models aside from others, that are solely obtained from fits to data (see Carling and Reader 1982, etc, labeled as "empirical predictors" in the introduction). This is justified since there is indeed a theoretical basis for these models, which is then intermixed with heuristics and fits to tackle the inherently challenging task. The same term is used in the work of Frings et al (2011), of which the present article can be seen as a continuation. We would thus argue that keeping the nomenclature is justified.

- 33 measurements → estimates (because a simulation is not a measurement)

  We changed it to the term "evaluations" to denote the porosity evaluation that is achieved for both, experiments and simulations.

- 63 between the largest and the smallest grains → between the diameter of the largest and the smallest grains

  Done.

- 185 binary packings → for binary packings of spheres

  While it is true that the parameterization of the models was most often obtained from sphere packings in the literature, that is not always the case (see CPM). Furthermore, the value of $n_{0,i} = 0.4$ was here chosen primarily for convenience and clarity in the plotting, and thus deviates from the value of 0.36 typically found for spheres. We adopted the text to clarify this.

- 244 for the here considered binary case → for the binary case considered here

  Done.

- 284 Additionally, and again as before, → As before, ...

  Done.

- 285 anew → new

  Integrated in following change.

- 285 Therefore, no anew adaption of the models to the validation data had been carried out here → Therefore, we did not have to adapt the models for the validation.

  Done.

- 299 here considered → considered here

  Done as "investigated here".

- 316 remove "as also used ..."

  Done, and reformulated.

- 375 The value of the initial porosity → The initial porosity

  Done.

- 393 for the size classes → between the size classes
  Done.
- 399 The following validation → The validation
  Done.
- 399 exactly equal → equal
  Done.
- 401 Here, it had to be adapted → Here, we had to adapt
  Done.
- 404 these laboratory measurements → the laboratory measurements
  Done.
- 404 The porosity prediction that we achieved with this combination → The porosity that we predicted with this combination
  Done.
- 405 measurements and simulation → measurements and simulations
  Done.
- 406 Such influence factors were fully captured by the applied measurement method in the laboratory but were not accounted for by the packing model, and also excluded in the simulations. → Such factors were fully captured by the laboratory measurement but were neither accounted for in the packing model nor the simulations.
  Done.
- 410 available → available.
  Done.
- 412 Rhine sediment → Rhine
  Done.
- 412 deviation from → difference to
  Done.
- 430 here considered grains → grains considered here
  Done.
- 444 here considered prediction models → prediction models considered here
  Done.
- 446 remove "In this paper,"
  Done.
- 446 presented and discussed → compared
  Done.
- 446 named → the
  Done.
- 457 This applicability was verified by 20 further simulations, →We verified the models further by comparing them to 20 simulations
  Done.
- 463 0.02mm → 0.02 mm
  Done.

- 467 input variables → input parameters
  Done.
- 472 theoretical packing models → why theoretical?
  We removed "theoretical".
- 472 Overall, we could showcase → Overall, we show
  Done.