Geochronology Discuss., author comment AC2
https://doi.org/10.5194/gchron-2021-39-AC2, 2022

Reply on RC1

Michael Dietze et al.

Author comment on "sandbox – creating and analysing synthetic sediment sections with R" by Michael Dietze et al., Geochronology Discuss., https://doi.org/10.5194/gchron-2021-39-AC2, 2022

We thank both referees for their assessment of the manuscript and the constructive and through provoking suggestions. We have implemented changes were needed and useful and explain in this letter where we have implemented changes and why. Please find below the comments and the corresponding replies.

Referee 1

1 Strengths

The lead author is one of the most accomplished R programmers in the Earth Sciences. The sandbox package is logically implemented and well documented. It follows R's list-based programming paradigm and uses the language's strengths in data visualisation whilst avoiding some of its weaknesses. For example, the package uses parallel processing for its calculation, thereby overcoming R's computational limitations. All this is achieved with minimal dependencies. Unlike many other packages, sandbox does not require ggplot2, which makes the package light and nimble.

Reply: We are thankful for the praise. Indeed, 'sandbox' is developed to run with minimal burden to be slim yet using mainly native R code. Yes, we decisively avoided chart junk overhead.

2 Major comments

2.1 Usefulness

I am not sure how useful sandbox is, and I doubt that it will ever become very popular. The three applications provided in Section 3.5 are not very convincing. Take, for example, the first case study, which investigates the effect of sample geometry on the single grain OSL age distribution. It shows that a cylindrical container results in a circular age distribution:

Reply: We made the examples decisively crisp and easy to follow in this introductory article. It is correct that sample container geometry effects (also in combination with different deposition rates) could be studied with less overhead code than required with ‘sandbox’. The main purpose of this example is to document the general way of using ‘sandbox’. There is always the chance to develop more elaborated, sophisticated and insightful examples. But when the first order goal is to let a reader follow the processing steps, then the deeper usefulness of the result may not necessarily dictate the example’s scope.

That said, we understand the general criticism and have now explicitly explained in the text why we use simple examples instead of elaborated ones pursuing a non-trivial research question (last paragraph of the introduction, first paragraph of section “Application examples and parameterisation”), and discuss that example 1 could also be solved more directly (first paragraph of discussion section “Geometric sampling effects”). Hence, we cannot estimate today, whether or not ‘sandbox’ will become very popular on the long run, however, for us the article is a start on which we can further build on.

The second application is similar to the first, whereas the third example did not make sense to me. I did not understand why sieving would cause ‘age inversions’ in the depth profiles. Do the coarse grains contain more inheritance than the small ones? If this is so, then I must have missed where this was specified. This should be explained better.

Reply: We have clarified the manuscript to better point at the underlying concept that we do indeed allow grains of different populations (hence size) to have a different inherited age and that their mixture with depth dependent population contributions can cause age inversions, even amplified through sieving if that sieving selectively enriches those populations with high inherited ages (section “Application examples and parameterisation”).

2.2 Crudeness

The virtual sediment sections defined by sandbox ‘rule books’ are not process-based. The algorithms are purely statistical and do not include any physics. This severely reduces their degree of realism, and limits their usefulness, as I will explain next.

Reply: This is statement correct, but intended by us. The package ‘sandbox’ is supposed to provide a general framework not a process-based sedimentation model. However, it allows to impose any physically based relationship and thus turn the model exercise into one that is constrained by physical “rules” of how the virtual section should be built and behave. We have clarified the manuscript and mention this concept now explicitly (first paragraph of section “Philosophy and structure of sandbox”).

Each location in the virtual sediment section is assumed to contain a discrete number of subpopulations. Each of these distributions is assumed to follow a lognormal grain size distribution with corresponding normally distributed mineral densities and grain packing densities. To generate virtual samples, random numbers are selected from these distributions. The grain size, mineral density and packing density are chosen independently, assuming zero covariance between their respective (log)normal populations. I think that this is an unjustifiable oversimplification. In real sedimentary sections, grain size, density and packing density are strongly correlated with each other.

For example, Stokes' Law dictates that small zircon grains (ρ = 4.65 g/cm3) are 'hydraulically equivalent' with larger quartz grains (ρ = 2.65 g/cm3). Therefore, well sorted sediment exhibits a size shift between quartz and zircon. In a sand that contains both quartz and zircon, the zircon will tend to fill the gaps between the sand grains, thereby increasing the packing density.

Reply: This is an excellent example to illustrate that 'sandbox' is not an all-case physically meaningful representation of sedimentation processes by default. It is correct that many grain (population) parameters are linked by chemical or physical properties and processes. However, as mentioned above, we decisively designed 'sandbox' to be like that. Meaning that by default 'sandbox' does not relate, for example grain size and specific density to packing density, does not mean that one cannot implement such a relationship.

We acknowledge this argument and have added several examples to the SI. There, we show how one could include specific density driven grain-size differences and the resulting packing density effects to a parameterisation of 'sandbox'. However, we think that including this rather elaborated example to the main manuscript would counteract our arguments made above (reply to point 2.1).

Although I am not an expert in OSL, I do think that this is important because zircon tends to be rich in actinides, and so the relationship between zircon and quartz affects the dose rate. Wouldn't this be a more important problem to simulate than the geometry of the sample container?

Reply: It is correct, the high concentration, in particular of U, in zircon grains might affect the dose rate to a level of which it may take over the role of a significant contributor causing strong dose-rate heterogeneities. However, compared, for instance to feldspar, representing a strong beta-emitter due to its potassium concentration, the effect is not of greater relevance in most sites where the amount of zircons is less than 1% of the overall sediment budget and usually well mixed within the sediment matrix. Moreover, if zircons are important they tend to accumulate in particular layers. In other words, while zircon related dose-rate effects on OSL ages may become more relevant than sample container shape effects in particular cases, the introduction of such a parameterisation example would inflate the manuscript. To compromise, we mention the zircon dose-rate effect as a motivation for a further potential application field of 'sandbox' (second paragraph of discussion section "Sample population effects").

In summary, hydraulic sorting and selective entrainment impose a strong covariance structure on the physical properties of sediments, which sandbox currently does not capture. In principle, it is possible to embed this covariance structure into the sandbox package using multivariate (log)normal distributions. However, in practice, this would not be so easy to implement, because it would dramatically increase the number of parameters that need to be set in the rule books. A process-based algorithm would fix this, but it would require a complete redesign of the package. Unfortunately, I can't think of a third solution.

Reply: We may argue for another solution. In the light of the correctly identified additional need to a) implement and b) parameterise one out of a considerable range of physical laws of sedimentation, which may be adequate in some but certainly not in all use cases, why not exemplarily show the basic steps of how to do this if needed and otherwise keep the flexibility of 'sandbox' to let users chose if and/or which physical relationships they want to use for their rule book? By following that latter suggestion, the effort of excessive parameterisation is separated from the 'sandbox' model definition and moved to external duties, as illustrated in the supplementary information.

In summary, we did the following to solve the raised issue: i) explicitly mentioning in the manuscript that realistic representations of sedimentary deposits may require the implementation of further physical relationships (second paragraph of section "Limitations"), and ii) work through its implementation in the supplementary information.

3 Other comments

3.1. It is not clear from the title why this paper was submitted to Geochronology. It is only in Section 3.5 (line 335) that the geochronological relevance of the sandbox package becomes apparent. This section uses the method to create some virtual OSL samples. This should be changed, especially because I am doubtful that sandbox will ever be used for any other applications. Suggestion: change the title of the paper to: "sandbox – Creating and Analysing Synthetic OSL samples with R."

Reply: We decisively picked Geochronology because the overarching theme of 'sandbox' is age-depth-parameter relationships of sedimentary deposits. This is a topic of paramount relevance for dating applications but often not exploited with the needed care. This becomes even more true when effects of the grain size structure are considered as systematic uncertainties that go into, for instance, Bayesian modelling frameworks, such as 'BayLum'.

The only true alternative to GChron (excluding non-open access journals) would be GMD. However, we are afraid that the circle of people we want to reach and inspire by 'sandbox' will not enthusiastically follow GMD as outlet of papers. When adding parameters like major or minor elements, cosmogenic nuclide concentrations as well as carbon fractions or other biomarker content of grain populations, we believe that 'sandbox' may well become an interesting tool for a wide range of scientists occupied with other than luminescence-dating routines.

3.2. All the random samples in the sandbox package are drawn from (log)normal distributions, which are extracted from real datasets using Dietze and Dietze's EMMAgeo package. An alternative and more flexible approach would be to draw random numbers from any cumulative distribution.

Reply: Parameters (and rules) can be defined using not just normal distribution functions, but also uniform, gamma and exact, as described in the manuscript and package documentation. We initially had the option to use empirical distribution functions, as well. However, the overhead and parameterize hassle was significant, so we decided (for now) to skip this additional option, assuming that gamma, normal, uniform and exact (no scatter) already provide quite a bit of freedom to define parameters. We should point out though that the grain-size definition indeed is in phi scale. This does limit flexibility, but only for this single parameter. However, that constraint was decisively made, in agreement with long standing support from sedimentology.

To clarify, EMMAgeo does not necessarily produce log-normal end-members but would in principal return any distributions underlying a mixed data set. In a recent paper (Dietze et al., 2021 DOI: 10.1111/sed.12929) this has been tested with synthetic data.

3. The paper is too long. sandbox is based on an inherently simple idea that I am confident could be explained in a paper half the length of the current manuscript. I

enjoyed reading the example code and the supplementary information item. However, I must confess that I found the main text a bit tedious to get through.

Reply: We shortened the original manuscript by 16 % (but had to add some requested text, yielding a total reduction by 12 %), a point also raised by referee two.