Response to comments posted by Referee #2

- Author's comment
- Comments from referee #2

We thank the second referee for reviewing our article and providing his/her feedback. We have addressed all of his/her comments and discussed them. The observations were partially complementary to referee #1 and very helpful to identify some unclear issues.

This paper presents a method called Histogram via Entropy Reduction (HER) for the interpolation of spatial geophysical data. This method is based on information theory measures of entropy and relative entropy, and has advantages over benchmarks of kriging and nearest neighbor methods in terms of its generality and lack of assumptions. The authors present the methodology which determines spatial dependency structure based on observed data points, and estimates optimal weighting parameters used to predict a given variable at a location. An application to several synthetic datasets shows high effectiveness of the method relative to three existing interpolation techniques.

Overall this paper was interesting and clearly written, and the figures are very informative and help to illustrate complex concepts. Although I do not have a background in geostatistics or interpolation of sparse datasets, this paper seems to introduce a promising avenue of how IT measures can be advantageous in this field. Some comments and suggestions listed below, which consist of minor revisions/technical corrections. They mainly highlight places that could use additional explanation or clarification.

Main Comments:

Comment 1: Line 111: on the description of creating the "infogram cloud": at first it was not clear to me whether an "infogram" was an existing technique that I was unfamiliar with, or designed by the authors. I think it is the latter (based on line 134) – either way, this aspect could be made more clear earlier in the subsection, that you have developed this graphical technique called an infogram that shows spatial correlation structure.

Response 1: Indeed, the term "infogram cloud" in L.111 is out of place. The same observation was made by referee #1. To avoid early questions, we propose rephrasing the sentence and including the name of the 3 assets used for the spatial characterization (Infogram cloud, Δz PMFs, and Infogram) when they first appear (L.112-121) as follows:

"*As shown in Fig. 1a, the spatial characterization phase aims to obtain: Δz probability mass functions (PMFs), where z is the variable under study; the behavior of entropy as a function of lag distance (which the authors denominate 'infogram'); and, finally, the correlation length (range). These outputs are outlined in Fig. 2 and attained in the following steps: i. Infogram cloud (Fig. 2a): [...] ii. Δz PMFs (Fig. 2b): [...] iii. Infogram (Fig. 2c)*".

Comment 2: Line 145: Could you add a bit more information on what effect/advantage this has? It seems like attributing a small probability to every category would make a larger difference to some types of distributions than to others.

Response 2: For the application of HER (mainly when using the log-linear aggregation method), it is desirable to assure that all bins of the distribution have a nonzero probability. This guarantees that there is always an intersection when aggregating PMFs. In this way, when the intersection between two PMFs happens only on the previously empty bins, the resulting PMF is a uniform distribution, i.e., the method effectively applies a maximum-entropy approach.

In addition, Darscheid et al. (2018) checked the impact of five alternatives for nonzero probability to a range of typical distributions (uniform, Dirac, normal, multimodal, and irregular) and concluded that, for the cases where no distribution is known a priori, three methods (including the one used in the paper) performed well across analyzed distributions. In order to add more information regarding the nonzero probability, we suggest rewriting the paragraph as follows:

" *[...] The bin size was defined based on Thiesen et al. (2018), by comparing the cross entropy (Hpq=H(p)+DKL(p||q)) between the full learning set and subsamples for various bin widths. The selected one shows a stabilization of the cross entropy for small sample sizes, meaning that the bin size is reasonable for small and large sample sizes and analyzed distribution shapes. For favoring comparability, the bins were kept the same for all applications and performance calculations.*

Additionally, to avoid distributions with empty bins which might make the PMF combination (presented in Sect. 2.3.1) unfeasible, as recommended by Darscheid et al. (2018), we assigned a small probability equivalent to the probability of a single point-pair count to all bins in the histogram after converting it to a PMF by normalization. This guarantees that there is always an intersection when aggregating PMFs, and that we obtain a uniform distribution (maximum-entropy) in case we multiply distributions where the overlap happens uniquely on the previously empty bins. Furthermore, as shown in the Darscheid et al. (2018) study, for the cases where no distribution is known a priori, adding one counter to each empty bin performed well across different distributions."

Comment 3: Section 2.3.1: For readers less familiar with aggregating probabilities towards a spatial context, I this description could benefit from some sort of illustration or simple example that shows the difference between the different pooling operators in Eqs 4-7. For example, show two measurement points D1 and D2 with a target location A somewhere between them, and show how the measures differ. This actually become more clear to me with the later discussion in Line 445 onward, so maybe some of these aspects could be brought forward earlier.

Response 3: The authors agree that an illustration of the aggregation methods could be beneficial for showing the practical meaning of each one of the aggregation options later explored in the application case. Since the example in the spatial context is given (without illustration) in L.185-201 and L.445-448, we believe that we could explore the practical implication of the methods in the end of the section 2.3.1. We estimate that it will increase the size of the paper in half page (figure plus brief explanation). A preview of the additional figure and explanation is shown below.

"The practical differences between the pooling operators used in this paper are illustrated in Fig. 3, where Fig. 3a introduces the two PMFs to be combined, and Figs. 3b,c,d show the resulting PMFs for Eqs. (5), (4), and (7), respectively. In Fig. 3b, we use unitary PMF weights, so that the multiplication of the bins (AND aggregation) leads to a simple intersection of PMFs weighted by the bin height. In Fig. 3c, we use equal weights to both PMFs, and the resulting distribution is the arithmetic average of the bin probabilities. Fig. 3d shows a log-linear aggregation of the two previous distributions (Figs. 3b,c). In all three cases, if the weight of one distribution is set to one and the other is set to zero (not shown), the resulting PMF would be equal to the distribution which receives all the weight. Specifically for Eq. (7), this means that the final distribution may result in a pure AND, Eq. (5), or pure OR aggregation, Eq. (4), as special cases."

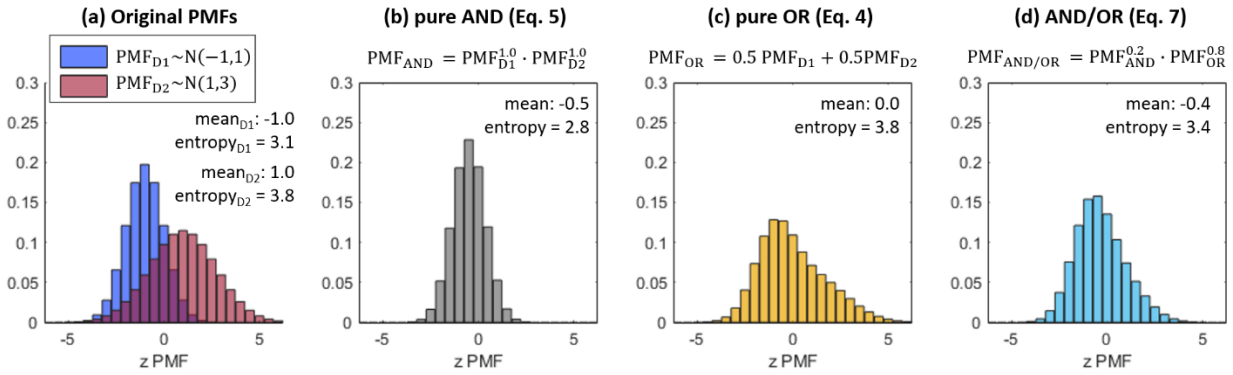

Figure 3: Examples of the different pooling operators. a) Normal PMFs N(μ,σ²) to be combined; b) log-linear aggregation, Eq. (5); c) linear aggregation, Eq. (4); and d) log-linear aggregation of (b) and (c), Eq. (7).

Comment 4: Section 3.1: Are there implications to model performance of the Gaussian process used to make the test cases? I wonder how this would compare to a realistic landscape (or whether this is considered very close to a "real world" case). Something is mentioned later about this in the discussion regarding the OK method, but more information would be beneficial here.

Response 4: Thanks for raising this question. The purpose of the paper is to test HER and demonstrate its performance in face of an established geostatistical method, namely OK. Thus, for testing HER, a field generated by Gaussian Process (GP) enables us to have a controlled dataset where we could examine their performance in fields with different characteristics (short and long range, with and without noise, small or large sample size). Since GP datasets fulfill the assumptions of Ordinary Kriging, it allows a fair comparison between the methods. We can say that GP and OK are the inverse of each other. While GP generates a dataset which follows a multivariate gaussian distribution with a known covariance function, OK estimates the stochastic process behind a dataset by fitting a variogram (or covariance function) and assuming that the residuals (i.e., estimation error) follow a gaussian distribution (Kitanidis, 1997, p.95).

It is true that a real-world dataset may not necessarily have the gaussianity properties given by the GP. Therefore it is the role of the geostatistician to guarantee that the data fulfill the method assumptions. . When it is not the case, e.g., it is common to transform the data so that it fits the assumptions, and back-transform it in the end. It is worth mentioning that, while developing the manuscript, the authors tested HER in a real-world case of digital elevation model data. Although HER and OK both performed well, its inclusion would require a proper geostatistical description of the dataset, which would be out of the scope of this paper and therefore we discarded its

presentation to keep the paper as short as possible and because GP covered a broader scope of fields. Altogether, this means that the test proposed using GP is also related to a real-world problem.
The authors understand that, due to being non-parametric, HER can deal with different data properties without the need of transforming the available data. HER does not require fitting of a theoretical function for extracting the spatial correlation, because its spatial dependence structure is derived directly from the available data. And since HER uses binned transformation of the data, it is also possible to handle binary (e.g. contaminated and safe areas) or even, with small adaptations, handle categorical data (soil types), covering another spectrum of real data.
Although parts of the above discussion were mentioned in L.470 and L.484, the authors believe that it would be beneficial to review the discussion section (Sect. 4.3) to include the above arguments in the manuscript.

Comment 5: Line 360: I think this is explaining why the infogram illustration in Figure 2a is very different from that in Figure 4a for the farther distances (which is because with the data, there are fewer pairs that exist at those farthest distances). If this is correctly interpreted, it could be brought up in the description of Figure 2a and the infogram cloud-shape in general.
Response 5: Thanks for pointing it out. As Fig. 2 as a whole merely illustrates what one could expect to get from the spatial characterization part, we decided to show basically the behavior of the first distance classes, where the spatial correlation is stronger. "Ignoring" the last classes is a common practice when analyzing spatial correlation, when the geostatistician defines a distance cutoff (maximum lag) for their analysis. Thus, considering that this omission, although discussed in Fig.4a, is in principle expected, the authors suggest including in L.129 the following clarification "*Note that in the illustrative case of Fig. 2, we limited the number of classes shown to four classes beyond the range. A complete infogram cloud and infogram is presented and discussed in the method application, Fig. 4.* "

Technical/writing style comments:

Comment 6: Abstract: There are several sentences here with parenthesis for additional context, I would recommend re-writing these without as many parentheses to potentially simplify and help the flow.
Response 6: Thanks. We will adapt the writing style in a revised version of the manuscript considering this point.

Comment 7: Line 38: applying
Response 7: Ok, thanks.

Comment 8: Line 90: H(X) is upper bounded by infinity in a continuous case, but as you mention in the next sentence and the equation that this case is discrete – the upper bound should be log2(N) where N is the number of bins or categories.
Response 8: We agree that it can cause misunderstanding, we will refine this paragraph.

Comment 9: Figure 6: It is hard to see the targets in a few of the maps, I think because the markers are in the background behind the observation markers. It would help to make outlines of the marker shapes bolder or colored to show which target is which.
Response 9: Thanks. We will endeavor to improve the visibility of the points.

Comment 10: Line 379: “differentiate between”, instead of differ?
Response 10: Ok, thanks.

Comment 11: Line 429: I found this sentence to be unecessary
Response 11: Thanks. We will consider removing it in the paper writing style revision.

References:

Darscheid, P., Guthke, A. and Ehret, U.: A maximum-entropy method to estimate discrete distributions from samples ensuring nonzero probabilities, Entropy, 20(8), 601, doi:10.3390/e20080601, 2018.

Kitanidis, P. K.: Introduction to geostatistics: applications in hydrogeology, Cambridge University Press, Cambridge, United Kingdom., 1997.

To the editor

We also suggest including in the discussion section (Sect. 4.3) two issues noticed by the authors during the revision process and while testing the method for assessing data uncertainty. The first one is that, since the dataset was evenly spaced, a possible issue of redundant information in case of clustered samples was not considered. Another matter that we wish to briefly discuss and propose theoretical solutions to is that, depending on how the first distance class is chosen, HER can lead to undesired uncertainty when predicting the value at the observations themselves.
In addition, we would also like to review the manuscript including some terminologies which are more precise to describe the method and its implications, and they could assist to foster the proper search for scholarly literature, mainly: E-type estimate (for the expected value obtained using HER) and conditional distribution (for the results of the aggregation method and PMFs obtained with HER). Although these terms are implicit in the method and explained, the authors would like to include them explicitly.