Karlsruhe Institute of Technology

Institute for Water and River Basin Management
Chair of Hydrology

HER: an information theoretic alternative for geostatistics

Stephanie Thiesen; Diego Vieira; Mirko Mälicke; Ralf Loritz; Florian Wellmann; Uwe Ehret

1 Introduction

The study proposes a stochastic, non-parametric, geostatistical estimator which combines measures of **information theory** with **probability aggregation method** for estimating the conditional distribution of a variable at an unsampled location. The method minimizes predictivity uncertainty, relaxes normality assumptions and therefore avoids the risk of adding information not present in the data. The approach is called histogram via entropy reduction (HER). We investigate the utility of our framework using different aggregation methods in a synthetically generated dataset from a known Gaussian process. We validate the efficacy of the method in ascertaining the underlying true field.

Example: Let's say we have measurements for the points A, B and C, and we want to predict the distribution of X.
With the spatial characterization step, we can associate the Δz distribution for each one of the points based on their distance class in relation to X, inferring P(X|A), P(X|B) and P(X|C).

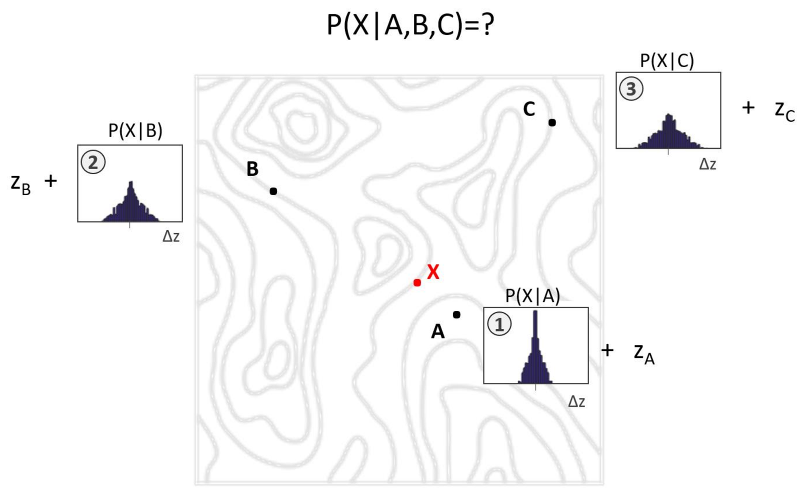

Figure 3: Schematic example of HER.

Finally, for obtaining P(X|A,B,C), we combine these conditional probabilities applying an aggregation method, e.g., linear pooling or log-linear pooling.

2 Method

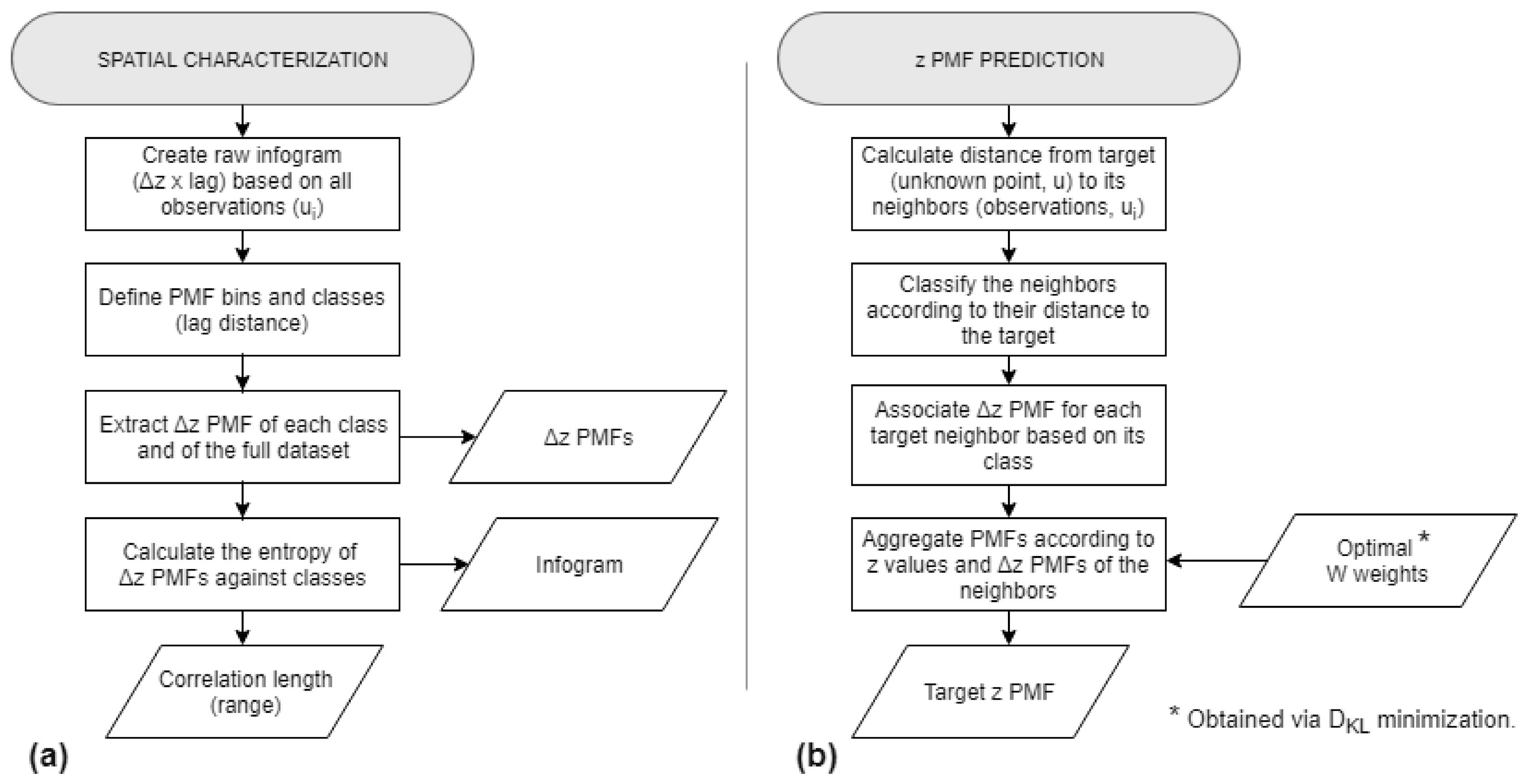

Figure 1: HER method. Flowcharts illustrating the a) spatial characterization, and b) probability mass functions (PMF) prediction.

Probability aggregation:

$$P(A|D_1,\dots,D_n) \approx P_G(P(A|D_1),\dots,P(A|D_n))$$

Log-linear pooling (AND)

$$P_{G_{AND}}(A) \propto \prod_{i=1}^{n} P_i(A)^{w_{AND_i}}$$

Linear pooling (OR)

$$P_{G_{OR}}(A) = \sum_{i=1}^{n} w_{OR_i} P_i(A)$$

Entropy:

$$H(X) = -\sum_{x\in\chi} p(x)\log_2 p(x)$$

Kullback-Leibler divergence:

$$D_{KL}(p||q) = \sum_{x\in\chi} p(x)\log_2 \frac{p(x)}{q(x)}$$

Predictions are directly based on empirical probability distributions, thus bypassing the usual steps of variogram fitting and normality assumption done in the traditional kriging method. In particular, the applied probability aggregation method provides a proper framework for uncertainty estimation that takes into account both the spatial configuration of the data as well as data values, while allowing to infer (or introduce) physical properties from the field under study.

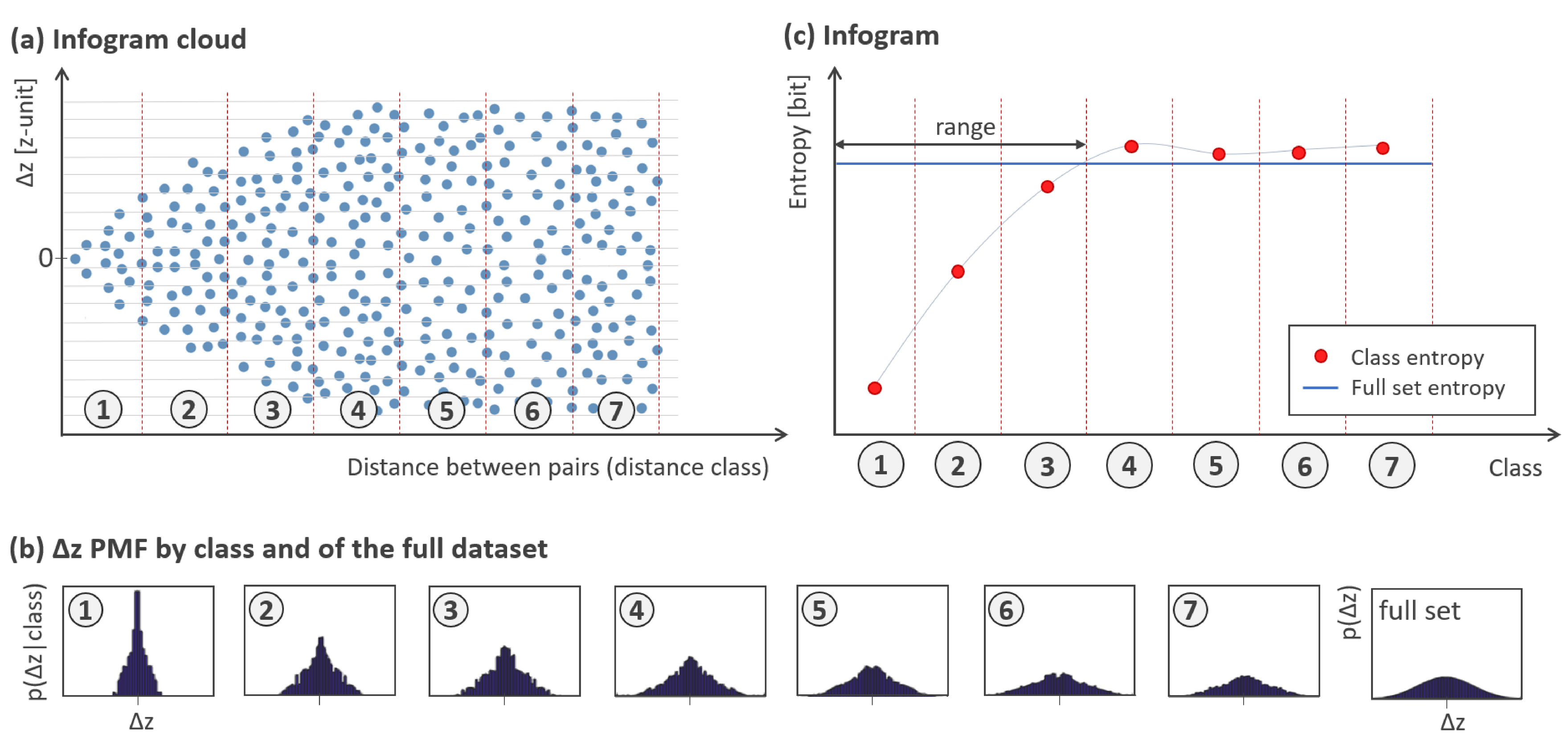

Figure 2: Spatial characterization. Illustration of the a) infogram cloud, b) Δz PMFs, and c) infogram.

3 Method application

This section presents three variants of HER models applied to the long-range field with noise (LR1) with a learning subset of 600 observations (LR1-600). The data was generated from a known Gaussian process, as specified in Figure 4.

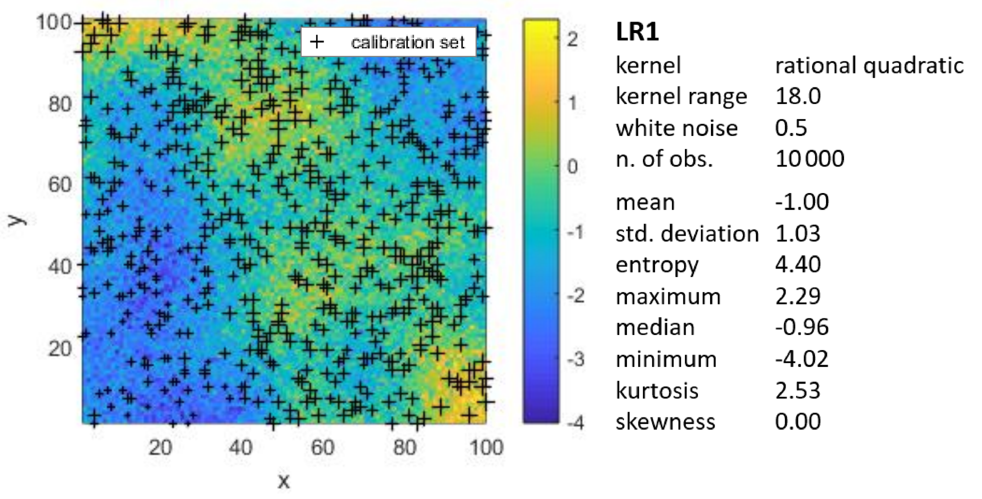

Figure 4: Synthetic fields and summary statistics.

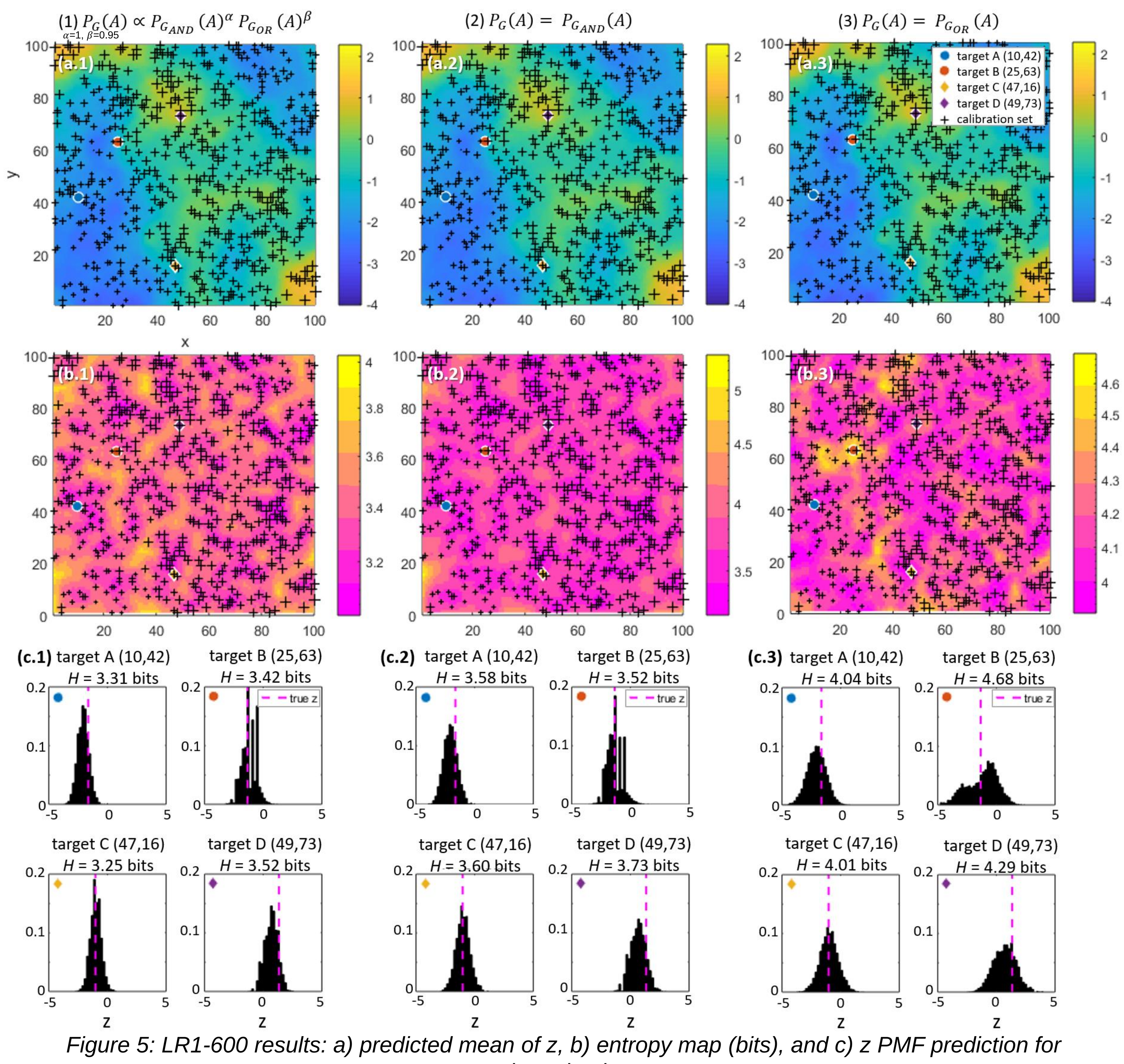

Figure 5: LR1-600 results: a) predicted mean of z, b) entropy map (bits), and c) z PMF prediction for selected points.

4 Conclusions

- Novel non-parametric, probabilistic method for geostatistics;
- Flexible framework for inferring or introducing physical properties from a field under study;
- By construction, conditional distribution are estimated, taking into account not just the spatial configuration of data but also the data values;
- Robust against fitting of spatial correlation function;
- Competitive performance compared to popular benchmark models, namely ordinary Kriging and inverse of distance (not shown);
- HER brings a new perspective of spatial and uncertainty analysis to geostatistics, using the lens of information theory;
- Additional investigation is required to analyze the method in the face of spatio-temporal domains, probability and uncertainties maps, adding co-variates.