Some theory for Multiple Point Statistics: Fitting, checking and optimally exploiting the training image

K.G. van den Boogaart[1]
1 Institute for Mathematics and Informatics, Greifswald University, Germany
Corresponding author: boogaart@uni-greifswald.de

ABSTRACT: In multiple point statistics, training images are used as a model for high order joint distributions. However this model does not necessarily represent the true joint distribution of the random field under consideration, like an arbitrarily guessed or estimated variogram does not necessarily represent the true correlation structure. This contribution thus tries to start something like variography for training images. However only first pictures fit on four pages.

KEYWORDS: multiple point statistics, training image, isotropy, selfsimilarity, scale, reproducibility, goodness of fit

1. Introduction

The Multiple Point Statistics approach of Strebel (2002) was introduced as an alternate approach to categorical geostatistics and uses instead of a variogram a user given training image to estimate the conditional probability at interpolation location given the observed and the already interpolated data.
The purpose of this contribution is to emphasis the conceptual proximity, similarities and differences of training images and variogram models and to show how well known concepts from variogram can be applied to the training image.

2. Training images replace variograms

According to Journel (2005) the training image is a conceptional model of the random process. However its more than that. It is also a quantitative model. Like specifying a variogram model the Specification of the training image determines the conditional expectation, conditional variance and conditional probability assumed by the multiple point statistics model and simulations. Fixing an ad-hoc training image is just like fixing a an ad hoc covariogram and an ad hoc mean for simple kriging. For a classical geostatistical model for binary images the parameters would be: nugget/microvariability, mean/sill, range/scale and variogram shape. In figure 1 the mean effect is exemplified. The character and portion of black in both simulations based on two qualitatively similar but quantitatively different training images is quite different and this would be

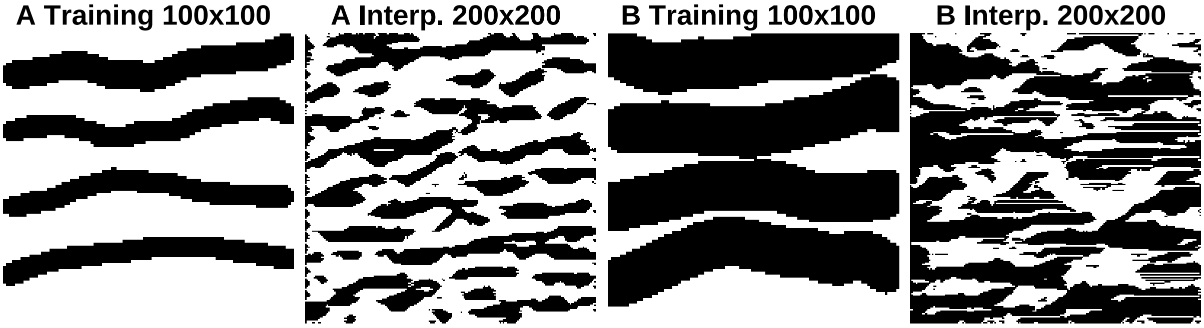

Figure 1: This figure shows two Multiple Points Statistics simulations based on the same dataset but different training images. The simulations are performed with the mps R-package from `http://www.stat.boogaart.de/mps`

the same for replica. A similar effect on the resulting image is illustrated in figure 2. With variogram the variogram model can be considered as the conceptual model of coregionalisation. However still the parameters of variograms are estimated from the data. To do similar things with training image we need have an idea how to parameterize a training image, which is picture drawn by an expert and we need an idea how to quantify the goodness of fit. Furthermore like understanding that a the empirical variogram does not give the true variogram we need to be clear that even the best training image does not give the true joint distribution of arbitrarily many points.

3. Parameterizing the training image

In our example the conceptual model is that of parallel channels. However without knowing the random field we can not know orientation, size, or portion of the channels. We will shortly discuss which parameters and decisions could be introduced

3.1. Scale as a range parameter

In figure 2 one can see the simple effect of scaling the training image with respect to the data. If we assume the natural process to be selfsimilar making all scales equal we can assume that the joint distribution comes from the superposition of all multivariate distributions from all scalings of the training image. The effect of this is shown in the rightmost figure of 2 and clearly induces a process having channels of various size. In opposite two point statistics of the variogram, where selfsimilarity already uniquely determines the form of the variogram we will have different self similar structures for different training images.

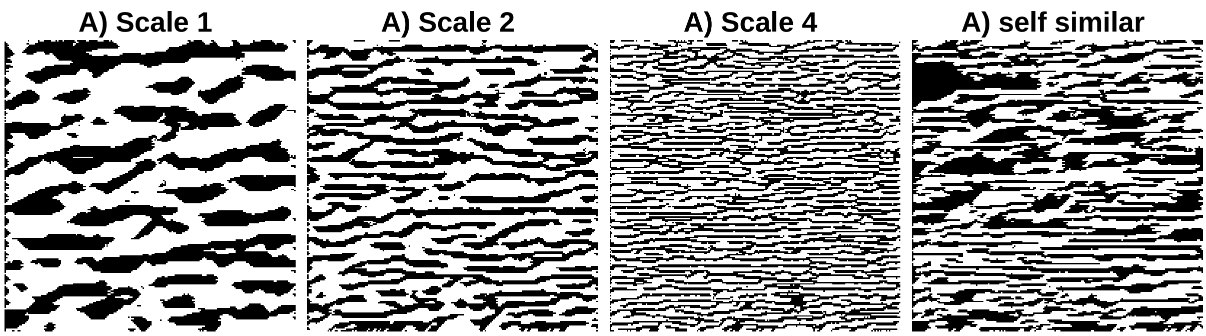

Figure 2: This figure shows two Multiple Points Statistics simulations based on the same dataset but with differently scaled training images. The last picture corresponds to the self-similar simulation proposed in this abstract.

3.2. Anisotropy direction and isotropy

Typically training images are not rotation invariant and thus give some axes of anisotropy which can be rotated. In our case we could even assume that the channels could have two typical directions. A typical assumption of geostatistics is isotropy, which can in this case realized by assuming that all rotations of the training image are equally valid. Since the whole procedure only uses local multivariate information from the training image the direction of anisotropy can vary locally in the simulation. The effects are illustrated in figure 3

3.3. The mean amount of the different categories

For binary images with finite range the sill functionally depends on the mean. So instead of varying the far range correlation we have to vary the relative amount of white and black for the conditional probabilities. The principle mathematical problem is to transform the marginals of the multivariate distributions by keeping their consistency. However already a simplified approach by just reweighting all entries of the multivariate contingency tables according to

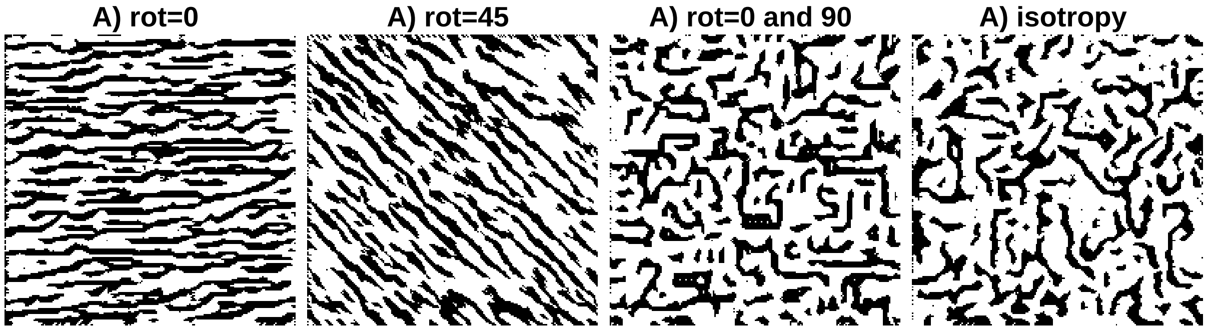

Figure 3: The effects of rotations and superposition of rotated training images on the interpolation. The third pictures clearly shows the interaction of two channel system. The last picture shows the effects of assuming isotropy leading to totally different quality of image but still preserving the idea of channels.

each of the marginals give a modified conditional probability changing the relative amount of black in the simulated images substantially.

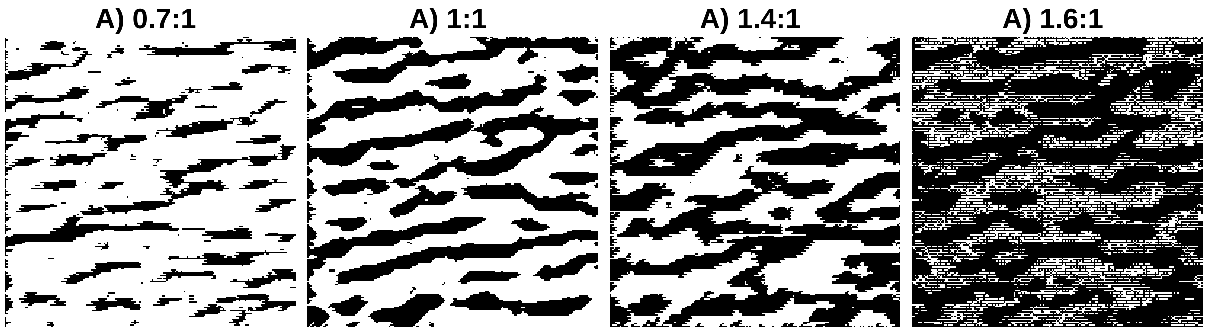

Figure 4: This shows the effects of reweighting the categories. While over a certain range the principle idea of channels is persevered the amount of channel in the total image changes substantially.

3.4. A nugget effect

A nugget effect will never be drawn in training image, like a nugget effect is not part of a variogram model but something additional. In the same way we can add certain portion of an independent marginal to the conditional distribution according to the training image. The effect is illustrated in figure 5.

4. Fit and reproducibility of the training image

The goodness of fit of a training image with respect to set of data can be quantified by the likelihood of local subconfigerations of observations with respect to their multivariate distribution according to the training image. We can use this measure to compare which combinations of parameters fits best.
Another aspect is that training image will never represent high order conditional distribution well enough, if the combinations are too seldom in the image is illustrated in figure 6. In the worst case the realized conditional probabilies are more or less a rough guess and would lead to an unrealistic distribution of the conditional simulations. More statistics is needed here.

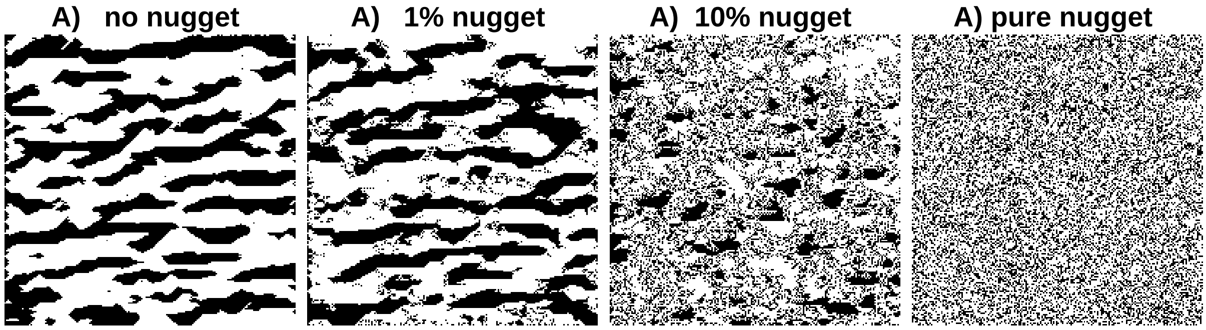

Figure 5: This shows the effect of adding a nugget effect by mixing the multivariate distribution with a certain amount of independent distribution up to total independency.

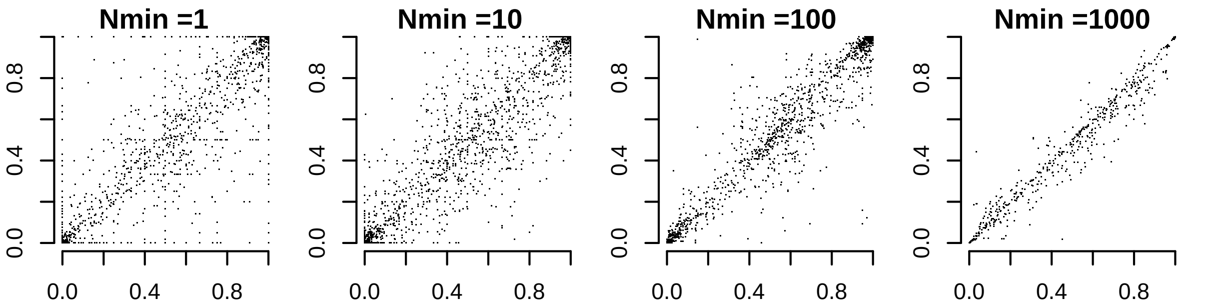

Figure 6: Reproducibility of conditional probability of "white" evaluated during the same MPS-run with respect the used on x and evaluated from alternate training image, which is a realization of the same random field, on y. From left to right an increasing number of minimum conditioning locations is used. If this could not be realized less conditioning sites are used. The leftmost picture represents what the classical algorithm does.

5. Conclusions

Although only one pictures was shown for each conditional simulation, the pictures shown should have made clear that the training image and the proposed parameters, have a strong effect on the distribution and the character of the conditionally simulated situations. A justification of the training image and the decisions on scale invariance or rotations invariance as well as fitting adequate parameters should be a mandatory task for the multiple point statistics approach in future.

REFERENCES:
Strebel, S. (2005) *Multiple Point statistics*, in Cheng, Qiuming, Graeme Bonham-Carter (Eds) (2005) Proceedings of IAMG'05, The Annual Conference of the International Association for Mathematical Geology, Toronto, ISBN 0-9734220-1-7
Journel, S. (2005) *Multiple Point statistics*, in Cheng, Qiuming, Graeme Bonham-Carter (Eds) (2005) Proceedings of IAMG'05, The Annual Conference of the International Association for Mathematical Geology, Toronto, ISBN 0-9734220-1-7