



Using Pattern Counts to Quantify the Difference Between a Pair of Three-Dimensional Realizations

Marie Lilleborge¹ · Ragnar Hauge¹ · Bjørn Fjellvoll¹ · Petter Abrahamsen¹

Received: 26 April 2023 / Accepted: 13 April 2024
© The Author(s) 2024

Abstract

When comparing different ways of modeling discrete three-dimensional realizations such as facies, it is useful to have a measure of difference (or similarity) in the geometry of these realizations. We propose a method for evaluating such difference by comparing pattern counts for a small template. Tests on synthetic datasets demonstrate that the proposed difference effectively differentiates between realizations of a Boolean model and those generated using multiple-point statistics with the Boolean realizations as training images. We also observed that multiple-point statistics realizations based on similar training images yield smaller differences to one another compared to those based on training images from dissimilar concepts. This suggests that the proposed difference is a useful tool for comparing discrete three-dimensional realizations.

Keywords Pattern count · Classification · 3D realization · Multiple-point statistics

1 Introduction

We propose a method to calculate a scalar difference between two three-dimensional gridded representations of categorical variables. The aim is to quantify the complex difference between the two representations into a single value. The challenge lies in capturing the essence of the difference within a single number that can evaluate significant geometric distinctions.

To demonstrate practical relevance, we use the proposed difference to evaluate the quality of simulation algorithms by comparing simulated realizations and quantifying their spread. Ideally, the difference between simulated realizations should be of the same order as the difference from a simulated realization to the corresponding training image representing the geological concept. We also use the difference to compare realizations generated from different geological concepts, including those derived from distinctly different training images.

✉ Marie Lilleborge
Lilleborge@nr.no

¹ Norwegian Computing Center, Oslo, Norway

The calculation of the difference between facies realizations was introduced in Park and Caers (2007), wherein a connectivity-based difference was used to efficiently explore various realizations in search of a good history match. Their proposed difference relied on time-of-flight between injectors and producers, with realizations with similar time-of-flight values essentially being assigned a low difference. Correspondingly, realizations with low difference are expected to exhibit similar production profiles. A notable aspect of Park's approach is its robustness to details in the realization that do not affect time-of-flight. Such details are likely of minimal significance in flow simulation modeling. However, this difference necessitates the presence of wells and depends on their locations.

Suzuki and Caers (2008) consider a Hausdorff distance. Like Park and Caers (2007), the objective was to use the calculated distance between realizations as a proxy for difference between production profiles. Unlike the time-of-flight-based difference, the Hausdorff distance in Suzuki and Caers (2008) can be computed without the need for wells. However, it is more sensitive to facies locations rather than their shapes. Consequently, it is suitable for determining similarity between a geological concept (training image) and one or more realizations. Various versions of the Hausdorff distance are discussed in Dubuisson and Jain (1994). Typically, these versions consider the largest difference between a point in one realization and a corresponding point in another, both representing the same facies. Accordingly, these types of difference emphasize the similarity in facies locations.

Implicitly, generative adversarial network (GAN) approaches (Zhang et al. 2019) also define a difference, as the GAN discriminator attempts to differentiate between generated realizations and the reference training image. However, this discriminator lacks transparency and focuses primarily on binary classification, specifically determining whether two realizations follow the same distribution, rather than quantifying the difference between them.

Our difference proposal is inspired by the multiple-point statistics (MPS) simulation algorithm (SNESIM) (Strebelle 2002). SNESIM aims to simulate facies realizations that replicate the pattern counts found in a template that scans a reference training image. Boisvert et al. (2010) proposed using the absolute difference between pattern densities in two images as the difference measure. In contrast, we propose a difference which assigns a small difference if the pattern counts are within random variation of each other, and a larger difference as the difference in pattern count increases. Both the difference proposed by Boisvert and the one proposed in this paper ensure that the calculated difference is robust to facies locations, but sensitive to facies shapes. Unlike Boisvert's proposed difference, ours incorporates a mechanism to discern between actual differences in distributions and random variation between realizations from the same distribution. Therefore, our difference proposal is well suited for determining whether two realizations are representative of the same pattern distribution. Modeling approaches like MPS and GAN are based on one or more reference training images. Ideally, these algorithms should produce simulated realizations that are indistinguishable from the training images.

In the following section, we describe the calculation of the scalar difference. In Sect. 3, we demonstrate an application of our proposed difference on realizations from a standard MPS model, testing its ability to distinguish between training images and

realizations for two different geological concepts. Finally, Sect. 4 provides a discussion and concluding remarks.

2 Evaluating the Difference

In this study, we consider patterns of a specific shape and size. Inspired by the MPS methodology, we examine all patterns within a finite three-dimensional template. Figure 1 illustrates a suitable template choice for many applications. By sliding the template across a realization, we identify the various patterns present, and count how many times each pattern appears.

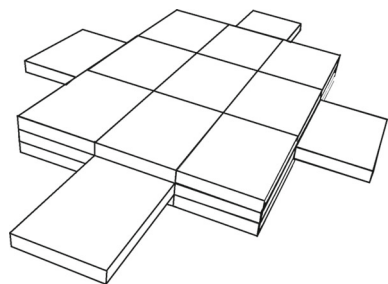
Consider a given pattern present n_1 times in one realization and n_2 times in another. Assume that the pattern is present at least 5 times in both realizations, that is, $n_1, n_2 \geq 5$. One aspect of the difference between the realizations is the degree to which the counts n_1 and n_2 differ.

Let us assume that the pattern counts follow a binomial distribution $\text{Bin}(N, p)$. Here, N represents the theoretical maximum count of a pattern determined by the template size and grid size of the realizations, while p is the success probability that may vary between patterns. The binomial distribution describes the number of successes (counts) in N independent trials. However, the assumption of independence is violated in this context for two reasons: Firstly, we count patterns in overlapping template locations, and secondly, there is spatial continuity in the rock facies in realizations and training images. Therefore, the assumption that pattern counts follow a binomial distribution is not valid in this context. Nevertheless, it remains useful for comparing pattern counts.

Consider the statistical hypothesis test with the null hypothesis that the pattern counts n_1 and n_2 arose from two binomial distributions $\text{Bin}(N, p_1)$ and $\text{Bin}(N, p_2)$ with equal success probabilities $H_0 : p_1 = p_2$. Let the alternative hypothesis be that the success probabilities differ, $H_1 : p_1 \neq p_2$. The two-sided p -value evaluates the extent to which the counts n_1 and n_2 differ for one pattern. Consider the test statistic

$$Z = \frac{|\hat{p}_1 - \hat{p}_2|}{\sqrt{2 \cdot \bar{p} \cdot (1 - \bar{p})/N}}$$

Fig. 1 The template used in this study had 31 cells over three layers (z -levels)



where $\hat{p}_1 = n_1/N$, $\hat{p}_2 = n_2/N$ and $\bar{p} = (\hat{p}_1 + \hat{p}_2)/2$. The test statistic Z has an approximate standard normal distribution under the null hypothesis H_0 whenever the proportion $\hat{p}_i \in [0, 1]$ is close to $1/2$. It has been determined that both counts n_1, n_2 should be at least 5 and at most $N - 5$ for the normal approximation to be valid (Campbell 2007).

A three-dimensional realization typically consists of various patterns, and an evaluation of the difference between two realizations should provide an overall assessment of whether the pattern counts differ between them. We propose the following three-step algorithm to achieve this:

1. Identify all patterns present at least 5 times in both realizations, and list them.
2. For each pattern in the list, calculate the two-sided p -value for its pattern counts n_1 and n_2 .
3. Report the proportion of p -values less than 0.05.

In summary, the proportion of p -values less than 0.05 serves as a summary statistic, representing the quantification of the difference between the two realizations. By using the p -value as an indicator of difference and removing patterns with very low occurrences, we obtain a robust difference across different scales of pattern frequencies.

3 Applications

We will demonstrate the ability of our method to distinguish between similar realizations and those generated from different statistical models. This study analyzes three-dimensional Boolean realizations of $400 \times 400 \times 50$ cells using a template of size 31. As template, we used cells within a Manhattan distance ≤ 2 from a center cell within the three layers directly above, at the same height, and directly below the center cell, respectively (Fig. 1). Then, the number of different template locations that fit within each realization is

$$N = (400 - 2 \cdot 2) \cdot (400 - 2 \cdot 2) \cdot (50 - 2 \cdot 1) = 7,527,168.$$

We expect this template to be suitable for many applications, as it covers a volume around the center cell without being unreasonably large.

3.1 Discrimination Between Models

We counted the frequency of each pattern in 20 three-dimensional realizations from a Boolean facies modeling algorithm (Holden et al. 1998). These were generated as 10 wide channels realizations and 10 narrow channels realizations (see Appendix 1). Two representative realizations are depicted in Fig. 2.

The six most prevalent patterns are visualized in Table 1. The prevalence of each pattern was similar within and across models, albeit with higher consistency within models. This is illustrated in Fig. 3, which presents two scatter plots of pattern counts from a pair of realizations generated by the same model (with wide and narrow chan-

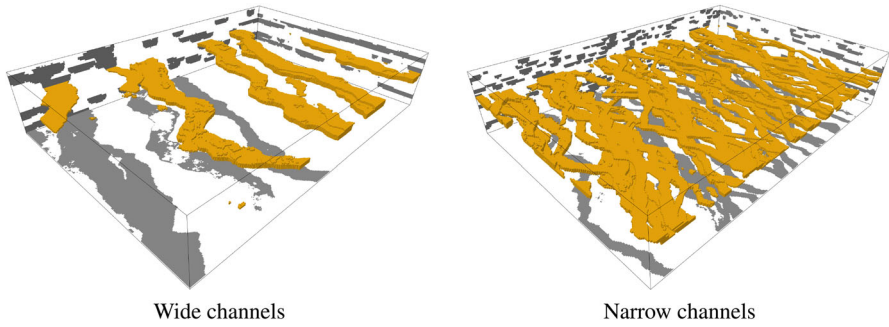


Fig. 2 Examples of one realization with wide and one realization with narrow channels generated using a Boolean object model (with the same volume fraction)

Table 1 Mean value (standard deviation) of pattern counts for the six most prevalent patterns in realizations with narrow channels and wide channels, respectively

	Pattern in template					
Narrow channels						
Mean pattern count	3,201,891	591,511	342,482	331,407	222,738	221,837
Standard deviation	82,970	17,540	9,435	6,276	8,310	7,736
Wide channels						
Mean pattern count	4,210,546	1,353,288	247,088	240,743	153,309	148,741
Standard deviation	52,584	33,170	8,877	10,754	3,585	4,875

Note that each of the most prevalent patterns had unique facies (gray: channel, white: background) across every layer of the template

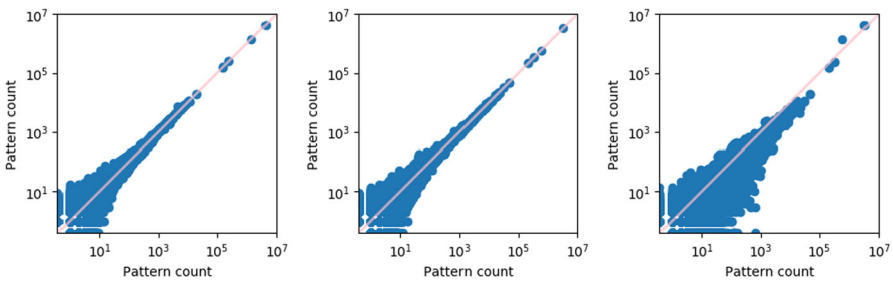


Fig. 3 Scatterplots of pattern counts within models (left panel: two realizations with wide channels, middle panel: two realizations with narrow channels) and across models (right panel: realization with wide compared to realization with narrow channels). Zero counts are represented at value 0.4 to appear on the log scale, with a slight gap to the nonzero counts

nels, respectively) and one scatter plot of pattern counts from two realizations from different models.

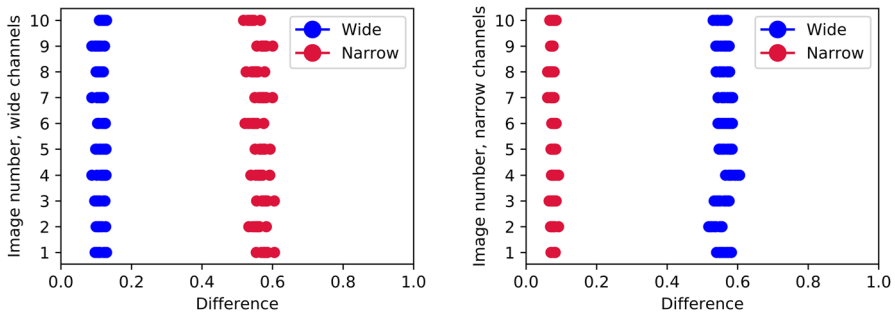


Fig. 4 Quantification of differences among all realizations from each of the wide (left panel) and narrow (right panel) channels realizations

Table 2 Mean (standard deviation) difference quantification between two realizations from the same (within) versus from different (across) models, and corresponding ANOVA hypothesis test results and critical value for the test statistic at $\alpha = 0.05$ level

Within model pair		Across model pair		ANOVA test statistic		
Difference	# of pairs	Difference	# of pairs	Critical F	Observed F	p -value
0.094 (0.020)	90	0.564 (0.018)	100	3.891	2.8×10^4	1.7×10^{-207}

Our difference evaluates realizations from the same model as being more similar than realizations from different models. This is visualized in Fig. 4, which displays all pairwise differences. With 10 realizations from each model, we calculated 100 cross-model differences and 90 within-model differences. One-way analysis of variance (ANOVA) confirmed a statistically significant discrepancy in the difference measure within versus across models (see Table 2).

Hence, our difference successfully distinguished between realizations from different models.

3.2 Classification of Realizations

The 20 channel realizations discussed in the previous section serve as training images for the RMS multiple-point (AspenTech 2022) based on SNESIM. We generate 10 realizations for each of the 20 training images, resulting in 220 realizations. In the following, we examine all realizations, including 2×10 training images (wide and narrow channels) and $2 \times 10 \times 10$ MPS realizations. Two realizations, generated using training images with wide and narrow channels, respectively, are shown in Fig. 5.

In line with the results of the previous section, we observed lower differences between a pair of realizations using training images from the same model (either wide or narrow channels) than those using different models. This is illustrated using density plots in Fig. 6.

At finer granularity, we observed a tendency for lower differences among MPS realizations of training images from the same model, particularly between realizations

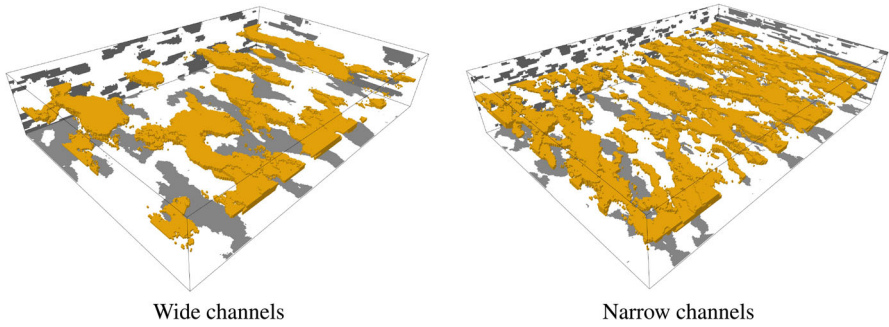


Fig. 5 Examples of one realization with wide and one realization with narrow channels generated using an MPS modeling algorithm. Compare to Fig. 2

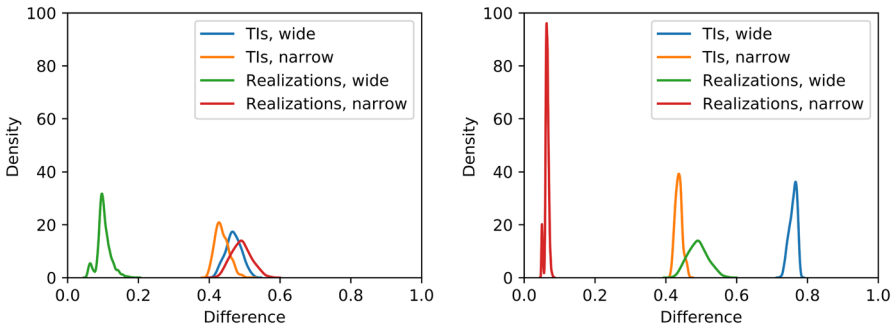


Fig. 6 Density of difference quantification from wide (left panel) and narrow (right panel) channel MPS realizations to all other realizations

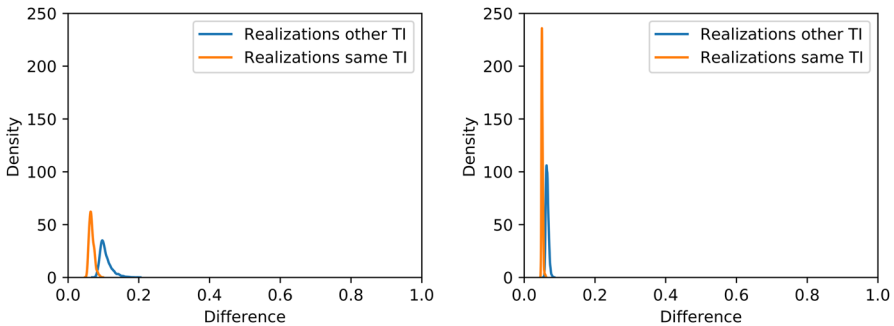


Fig. 7 Density of difference quantification for MPS realizations produced with training images from the same model, within wide (left panel) and within narrow (right panel) channel MPS realizations

from identical training images compared to those from identically distributed training images (see Fig. 7).

One-way ANOVA verified the visual observations (Table 3).

Finally, a two-dimensional multidimensional scaling plot provides a visualization of how the difference can be used to categorize the realizations into the four groups of training images and realizations with narrow and wide channels, respectively (see

Table 3 Mean (standard deviation) difference quantification between two realizations of the same type versus different types, along with corresponding ANOVA hypothesis test results and critical value for the test statistic at $\alpha = 0.01$ level

2 wide or 2 narrow		1 wide and 1 narrow		ANOVA test statistic		
Difference	# of pairs	Difference	# of pairs	Critical F	Observed F	p -value
Among all MPS realizations						
0.083 (0.024)	9,900	0.491 (0.029)	10,000	6.636	1.2×10^6	0.0
2 MPS same TI		2 MPS different TI		ANOVA test statistic		
Difference	# of pairs	Difference	# of pairs	Critical F	Observed F	p -value
Among MPS realizations with narrow channels						
0.051 (0.002)	450	0.065 (0.004)	4,500	6.640	6.0×10^3	0.0
Among MPS realizations with wide channels						
0.067 (0.007)	450	0.106 (0.016)	4,500	6.640	2.6×10^3	0.0
2 MPS or 2 TI		1 MPS and 1 TI		ANOVA test statistic		
Difference	# of pairs	Difference	# of pairs	Critical F	Observed F	p -value
Among all realizations with narrow channels						
0.064 (0.006)	4,995	0.436 (0.010)	1,000	6.639	2.7×10^4	0.0
Among all realizations with wide channels						
0.103 (0.019)	4,995	0.469 (0.022)	1,000	6.639	2.8×10^4	0.0

Fig. 8). All pairwise distances were set to the difference value between the realizations minus the expected false-positive rate (0.05), with a minimum distance of zero.

4 Discussion and Conclusions

In this study, our difference quantification based on pattern counts consistently distinguished between groups of realizations, particularly when the groups represented different models. The difference also demonstrated its ability to discriminate between various groups which were constructed to represent the same model, such as realizations generated from different training images of the same model, and training images compared to their realizations. However, realizations from the same model remained more similar than those from different models, even when generated from different training images of the same model.

The difference between MPS realizations and their corresponding training images was notably much larger than the difference between two MPS realizations from the same training image. Surprisingly, we also observed that realizations from different training images of the same model exhibit greater similarity to each other than their respective training images. This suggests that our difference can detect a common perturbation in the MPS realizations stemming from the same channel regime (wide or narrow), regardless of the training image used. One possible explanation for this

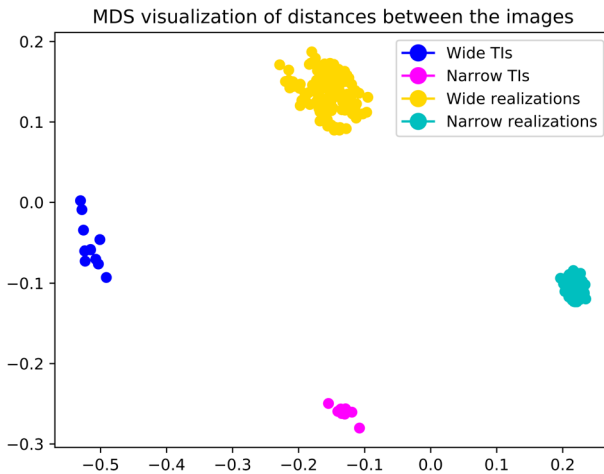


Fig. 8 Visualization of within-collection and across-collection difference plot for realizations with wide and narrow channels and training images. The two-dimensional multidimensional scaling plot was constructed using `sklearn.manifold.MDS` (Scikit 2023) with a maximum of 50,000 iterations and 100 different initializations

could be linked to the handling of scenarios where no legal patterns are identified during the simulation.

We do not believe these observations to be sensitive to the geometry of the template, unless compared to a template with a much larger or smaller number of cells: A minimal template comprising just a couple of cells would lack discriminatory power to distinguish between three-dimensional patterns, while a substantially larger template would produce an enormous variety of patterns with drastically lower pattern counts and unstable frequency estimates as a result. Given our observation that the template can discern between MPS simulations and their training images, we contend that its geometry is well suited for our intended purpose. We have not tested other templates for this paper. Our method could also be applied to multiple-facies cases. In such scenarios, it is reasonable to anticipate that more computing resources and larger training data would be needed.

We conclude that our difference enabled analyses capable of distinguishing between classes of images, including realizations derived from diverse training images. Furthermore, we observed that the difference assigned relatively small values between pairs of realizations generated from varied training images within the same model, compared to pairs of realizations where the training images originated from different models.

Appendix A: Input Parameters to the Boolean RMS Facies Algorithm

Realizations were simulated on a grid with 400 columns (Min X : 0.0–Max X : 19,199.3), 400 rows (Min Y : 0.0–Max Y : 13,326.5) and 50 layers (Min Z : 0.0–Max Z : 170.0) in a single zone with no faults. The rotation angle was 0.00 degrees clockwise, with rotation origin X of 24.0, rotation origin Y of 16.5. We used two distinct models of channels defined in the RMS facies modeling algorithm 'Channels NGOM' (AspenTech 2022). Both models used a volume fraction of 0.30 ± 0.02 , a channel orientation defined by azimuth (10, 2.5) and dip (0, 0.01) and channel wavelength 3,830 and amplitude (630, 63). The correlation between channel thickness and channel width was set to 0.5. Ten training images of narrow channels were generated with input parameters width (520, 25) and thickness (10, 1.5), and 10 training images of wide channels were generated with input parameters width (1,300, 65) and thickness (20, 4).

Acknowledgements This work was supported by the Research Council of Norway. The authors would like to thank Ingrid Aarnes for assistance with prettification of Figs. 2 and 5, and Hanna Marie Nicholas for line editing the manuscript.

Declarations

Conflict of interest The authors declare no Conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- AspenTech (2022) Smart reservoir modeling: Aspen RMS™ (Brochure). <https://www.aspentech.com/en/resources/brochure/smart-reservoir-modeling-aspen-rms>
- Boisvert JB, Pycrz MJ, Deutsch CV (2010) Multiple point metrics to assess categorical variable models. *Nat Resour Res* 19(3):165–175
- Campbell I (2007) Chi-squared and Fisher–Irwin tests of two-by-two tables with small sample recommendations. *Stat Med* 26(19):3661–3675
- Dubuisson MP, Jain A (1994) A modified Hausdorff distance for object matching. In: Proceedings of 12th international conference on pattern recognition, vol 1, pp 566–568
- Holden L, Hauge R, Skare Ø, Skorstad A (1998) Modeling of fluvial reservoirs with object models. *Math Geol* 30:473–496
- Park K, Caers J (2007) History matching in low-dimensional connectivity-vector space. In: EAGE conference on petroleum geostatistics, Sep 2007. European Association of Geoscientists and Engineers
- Scikit learn (2023) sklearn.manifold.mds. <https://scikit-learn.org/stable/modules/generated/sklearn.manifold.MDS.html>
- Strebelle S (2002) Conditional simulation of complex geological structures using multiple-point statistics. *Math Geol* 34:1–21

- Suzuki S, Caers J (2008) A distance-based prior model parameterization for constraining solutions of spatial inverse problems. *Math Geosci* 40:445–469
- Zhang T, Tilke P, Dupont E, Zhu L, Liang L, Bailey W (2019) Generating geologically realistic 3D reservoir facies models using deep learning of sedimentary architecture with generative adversarial networks. *Pet Sci* 16:541–549