

Assessing and Comparing Training Images

Jiaju Zeng and Jeff B. Boisvert

In order to have an effective application of multiple-point statistics simulation, an appropriate training image (TI) should be selected. The TI should match the expected spatial structure and geological features believed to be present in the deposit because the training image controls the spatial continuity and ordering relationships that will be simulated. Reservoir modeling professionals who have a significant level of geological knowledge can narrow the list of potential training images, but this is time consuming and subjective. This leads to the classification of all training images into smaller groups. Multiple-point statistics such as the distribution of runs and the multiple-point histogram from the exploration dataset are compared to the training images. The scale of the differences in the multiple-point statistics indicates the similarity between the datasets and the training images and helps rank potential training images. Experts can then examine the reduced number of training images and apply professional judgement to narrow the list of images and finally pick the appropriate training image. The choice of statistic will directly impact which TI is selected. The purpose of performing these calculations is to supplement but not remove the need for expert knowledge in TI selection.

Introduction

Multiple-Point Statistics (MPS) utilizes the influence of multiple separately located points to predict the local conditional distribution based on the training image (TI) statistics. Local conditioning data given is very limited and cannot provide a comprehensive understanding of the geological features and shapes. Using an analogue for the deposit to constrain the statistics is done through the selection of an appropriate TI to use in MPS.

The size of the TI must be large enough to include a number of repetitions of features of interest for the calculation of the MPS. Moreover, the TI should have the same properties of interest but may not originate from areas that have the same depositional system as the model area. Different areas have different specific properties and the TI must reflect the correct properties of interest. Once a set of realizations has been generated, the problem becomes one of TI selection. Ideally, an expert with significant knowledge of the domain would select an appropriate TI, but there are often multiple candidate TIs that would be appropriate. A method that attempts to classify TIs into groups with different geological properties will be introduced in this paper so that the volume of TIs reviewed is reduced as exemplars from different groups can be viewed. There are several types of MPS such as the connectivity function, lacunarity, the multiple-point density function (MPDF) and the distribution of runs are used to select the right TIs. The distribution of runs will be discussed and compared as the MPS of the TIs in order to split the TIs into groups. The TI will be 'drilled', i.e. samples extracted from the images and the statistics of the samples will be used to determine the appropriate TI to use. Ideally, the TI that the sample originated from will be selected.

Background

First and second-order statistics such as the histogram and the semivariogram are often utilized in traditional two-point geostatistics. Two-point geostatistics cannot capture the geological shapes or curvilinear structures present in real deposits, only linear spatial relationships are captured. Initial work in MPS by (Deutsch, 1992; Guardiano and Srivastava, 1993) lead to the consideration of MPS in modeling higher-order statistics inferred from a TI. This has the benefit of reproducing complicated geological structures and was made practical in SNESIM (Strebelle, 2002). Implementation of MPS is common and the interested reader is referred to the following for practical implementation of MPS in current workflows (Strebelle, Payrazyan, and Caers, 2002; Caers, Strebelle, and Payrazyan, 2003; Harding and others, 2004; Liu and others, 2004). The method discussed in this paper uses the distribution of runs and the multiple-point histogram to rank the potential TIs.

The Distribution of Runs

The distribution of runs (Mood, 1940) is a type of MPS that can be applied to a one dimensional string of values. If given data has continuous values, the values have to be classified into different categories by applying thresholds.

The concept of the distribution of runs is to count the number of consecutive occurrences of a particular category. For example, the distribution of runs for this example (Figure 1) would be 3 runs of length one, 2 runs of length two, and 1 run of length three.

The Multiple-Point Histogram

The multiple-point histogram (MPH) determines the number of possible unique multiple point configurations involved in a template specified by the user and shows the frequency of times each multiple point configuration occurs in an image (i.e. a 5 point one dimensional configuration that could take two different categories, there are a total of $2^5=32$ unique configurations to be counted).

For example, in a 2x2 multiple-point template with two categories the possible unique multiple-point configurations are shown in Figure 2. MPH counts the number of times each of these configurations occurs in an example image (Figure 3) and can be shown (Figure 4). Considering a larger template with more categories significantly increases the number of bins in the MP histogram and increases the difficulty of visualizing the MPS histogram. Therefore, the distribution of runs has an advantage over the MPH with larger number of categories involved in the data.

Accessing the Training Images

In order to determine the MPS such as the distribution of runs and multiple-point histogram of the TIs, the data values have to be binned into different categories if they are originally continuous variables. In some cases, continuous data values are used directly as TIs, but in this case we are interested in categories. Threshold limits are determined visually but should be selected by professionals who could distinguish the difference of the properties of different minerals. After the thresholds are applied a categorical model is generated for use as a TI.

Methodology

First step is to generate the categorical models of the TIs by applying threshold limits to the continuous data. The drillhole data is classified into categories with the same principle. The absolute difference between the MPS such as the distribution of runs and multiple-point histogram of the TIs and the available exploration data are compared and ranked. For each particular drillhole data, TIs are then ranked according to these differences (the smaller the difference, the higher the rank). Only MPS calculated in one dimension (along the drillhole) are considered due to the widely spaced vertical well data. As a result, any MPS used to compare TIs and exploration data are only informed in the vertical direction.

Select TIs with Distribution of Runs

The distribution of runs of each drillhole data set and the TIs are calculated independently. Once the results are obtained, the distribution of runs of the drillhole data can then be compared to the distribution of runs of all potential TIs. The comparison is based on the absolute difference between the distributions and summing over all run lengths.

Formula for calculating the difference between distributions is shown:

$$\Delta = \sum_{l=1}^n |f_l^* - f_l^{**}|$$

where Δ is the difference between the distributions, l is the length of runs, n is the maximum length of runs, f_l^* is the cumulative frequency of the well data for a run of length l , and f_l^{**} is the cumulative frequency of the TI for a run of length l (Boisvert et al, 2008).

Select TIs with Multiple-Point Histogram

A method which is identical to using the distribution of runs is used with the multiple-point histogram to compare each potential TIs to the drillhole data but the difference is now taken as the sum of the absolute difference between the frequency of each multiple-point bin.

Formula for calculating the difference between frequencies is shown:

$$\Delta = \sum_{i=1}^n |g_i^* - g_i^{**}|$$

where Δ is the difference between the frequencies, i is the bin number, n is the maximum number of bins, g_i^* is the frequency corresponding to bin i for the wells, and g_i^{**} is the frequency corresponding to bin i for the TI (Boisvert et al 2008). Again, the smaller is Δ , the more alike are the two multiple-point distributions.

It is important note that the use of a non-linear multiple-point configuration would not be possible because the exploration data is available in 1-D. Therefore, when using the MPhist to compare drillhole data to the library of potential TIs, multiple-point templates should all be 1-D. Figure 5 is an example of some three-point vertical 1-D multiple-point templates.

Case Study

A number of high resolution photos were taken in a granite warehouse. For every type of granite, two to three photos with different scales are available. After filtering out the bad photos (blurry, over reflective and etc.), a total of 26 high resolution photos are selected. In order to have photos with more varieties and different geological setting to prevent bias in the analysis, a number of photos of different rocks and outcrops are borrowed from the Saskatchewan Geological Survey. Nine representative photos are selected visually. A total of 35 photos are used to demonstrate the approach for narrowing down the set of potential TIs and finally selecting the appropriate TI. However, other useful statistics, such as the histogram and semivariogram, should also be used when selecting a TI; this is commonly done and is not considered here.

These high resolution digital photos are then converted into an array containing the image data. Since the photos contain true color image, the size of the array is M-by-N-by-3, where M and N are the sizes of the original photo in vertical and horizontal direction. Each pixel of the photo has a true color assigned therefore each slot of the array contain three individual values, red, green and blue values. Histograms of each color value can be calculated individually. To distinguish the spatial relationship between different minerals on the rock, thresholds in the R,G,B are introduced to provide a clear cut-off in between minerals which have a unique true color. An example subplot of the original photo, categorical model, R, G, B and gray scale histogram, and pixelplots of different color values is shown in Figure 6.

The selection of the threshold values is based on the R, G, B histogram. For example, in the red color histogram shown in Figure 6, two peak values are at 50 and 175, also a dip at 100 meaning the two majorities of the red color values are either below 100 or above 100. Therefore, 100 is taken as the threshold in red. There are two ways to divide the pixels into two categories 1) based on the R, G, B values, a combination of thresholds in each of the color values can define the categories 2) using the grayscale values as a fixed function of the R, G, B values:

$$\text{Gray} = R * 0.299 + G * 0.587 + B * 0.114$$

A single grayscale threshold can be applied to define the categories. In this case study, the latter method is used to create the categorical models. At this point, the array containing coordinates and R, G, B values is transformed into an array containing coordinates and category names (in this case, 1 and 2). These arrays are the categorical models of the Tis required for comparison. Three vertical mock drillholes data are then extracted from the each categorical model at a spacing of 800 pixels in the horizontal direction.

With all the TIs and the data sets ready in the form of categorical models, the distribution of runs can be calculated. An existing program is used to read in the results of the distribution of runs of each TI and drillhole and

compare. The absolute differences for all the comparisons are recorded. Also, the ranking of the matching TI for each drillhole dataset are stored in a separate file. Plots of the absolute differences are shown in Figure 7. Based on the calculations and these visualizations, the first 21 TIs after reordering will be classified into Group 1 and the rest into Group 2.

Figure 8 shows a good example on the left and relatively poor example on the right. The poor result, DS#35 does not have a spatial feature that can be distinguished clearly between two groups. Therefore, some TIs from both groups are ranked high. TI#35 is not similar to its sample holes as it is ranked 10th. However, 3/5 of the TIs in the top 5 are from the same TI group, which is an acceptable result.

The multiple-point histogram of each TI and DS are calculated (Figure 9). The results are not as decisive as the distribution of runs. In this example, the three drill holes do not provide sufficient information for an accurate comparison based on the multiple point histogram and the results are erratic.

In order to compare the functionality of the two statistics, it is useful to know how well correlated the differences between TIs are to the difference between DSs. For example, if the absolute difference from comparing TI#1 to TI#2 is high, it is expected that the absolute difference from comparing DS#1 to DS#2 is also relatively high. Figure 10 shows results of these two MPS are well correlated since both factors of correlation are greater than 0.6. However, the distribution of runs did a better job matching the DS differences to the TI differences. Histograms in Figure 11 indicate that method with distribution of runs has a better ability matching the DSs to the original TIs while MPhist is likely to rank the original TIs after the 20th. The distribution of runs seems to better identify the correct source TI than the MPhist.

Conclusion

In order to narrow down the list of the potential TIs and select the appropriate TI, MPS such as the distribution of runs and MPhist discussed in this paper may be applied for the comparison of exploration data with potential TIs. The choice of the MPS should be tailored to the setting of the exploration data and the geological features of interested. The comparisons between MPS from the available data sets and the TIs are used to supplement, not replace, expert geological knowledge. Although performing simple calculations of the difference between the TI and the data available can provide important information, expert knowledge should not be ignored.

Acknowledgement

Special thanks to Saskatchewan Geological Survey and Omni Surfaces Corporation for providing high resolution digital pictures for the case study.

References

- Boisvert, J.B., Pyrcz, M.J. and Deutsch, C.V., 2008, Multiple-Point Statistics for Training Image Selection; *Natural Resources Research*, Vol. 16, No.4, 313. ①
- Caers, J., 2001, Geostatistical reservoir modeling using statistical pattern recognition: *J. Petrol. Sci. Eng.*, v. 29, no. 3, p. 177–188.
- Deutsch, C. V., 1992, Annealing techniques applied to reservoir modeling and the integration of geological and engineering (Well Test) data: Stanford University, Stanford, 306 unpublished doctoral dissertation.
- Guardiano, F., and Srivastava M., 1993, Multivariate geostatistics, beyond bivariate moments, in Soares, A., ed., *Geostatistics Troia 1992*, v. 1, p. 133–144.
- Harding, A., Strebelle, S., Levy, M., Thorne, J., Xie, D., Leigh, S., Preece, R., and Scamman, R., 2004, Reservoir facies modeling: New advances in MPS, in Leunangthong, O., and Deutsch, C. V., eds., *Proceedings of the Seventh International Geostatistics Congress*, Banff, Alberta, 10 p.
- Hu, L.Y. and Chugunova, T., 2008, Multiple-point geostatistics for modeling subsurface heterogeneity: A comprehensive review, *Water Resour. Res.*, 44, W11413, doi:10.1029/2008WR006993.

Liu, Y., Harding, A., Abriel, W., and Strebelle, S., 2004, Multiplepoint statistics simulation integrating wells, seismic data and geology:Am. Assoc. Petrol. Geol. Bull., v. 88, no. 7, p. 905–921.

Liu, Y., 2006, Using the Snesim program for multiple-point statistical simulation; Computers & Geosciences, Vol. 32, Issue 10, p. 1544–1563.

Mood, A. M., 1940, The distribution theory of runs: Annal. Math. Stat., v. 11, no. 4, p. 367–392.

Pyrzcz, M. J., and Strebelle, S., 2006, Event-based Geostatistical modeling of deepwater systems, reservoir characterization: Integrating technology and business practices: Gulf coast section: SEPM Twenty-Sixth Annual Research Conference, p. 893–922.

Strebelle, S., Payrazyan, K., and Caers, J., 2002, Modeling of a deepwater turbidite reservoir conditional to seismic data using multiple-point geostatistics, in SPE Annual Conference and Technical Exhibition San Antonio, Soc. Petroleum Engineers Paper 77425, 10 p.

Ortiz, J. M., and Deutsch, C. V., 2004, Indicator simulation accounting for multiple point statistics: Math. Geol., v. 36, no. 5, p. 545–565.

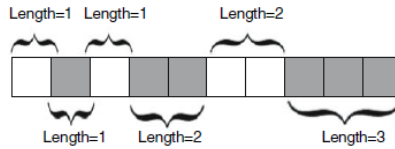


Figure 1. Calculating the distribution of runs for a well with two values, white and gray (Boisvert et al 2008)..

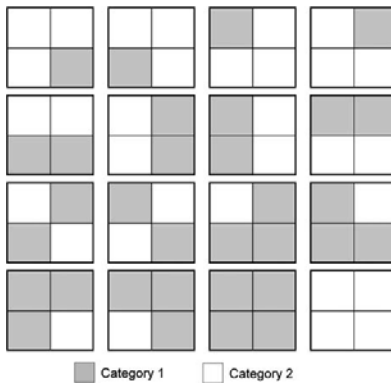


Figure 2. All possible configurations of a square four-point template with two categories (Boisvert et al 2008).

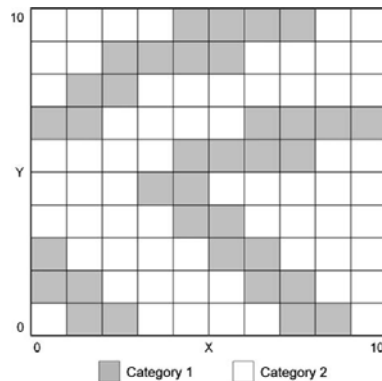


Figure 3. Example image to calculate the multiple-point histogram. The unit length is 10x10 cells (Boisvert et al 2008).

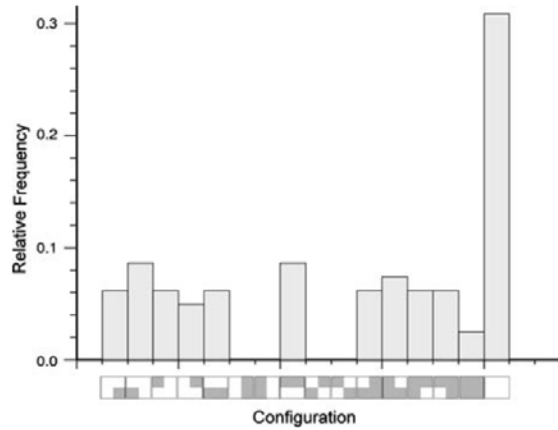


Figure 4. Multiple-point histogram of the example image shown in Figure 3. Each bin indicates a unique multiple-point configuration and the bin ordering has no significance (Boisvert et al 2008).

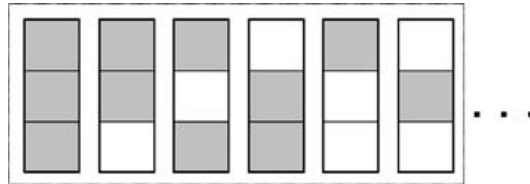


Figure 5. Example of some three-point vertical 1-D multiple-point templates (Boisvert et al 2008).

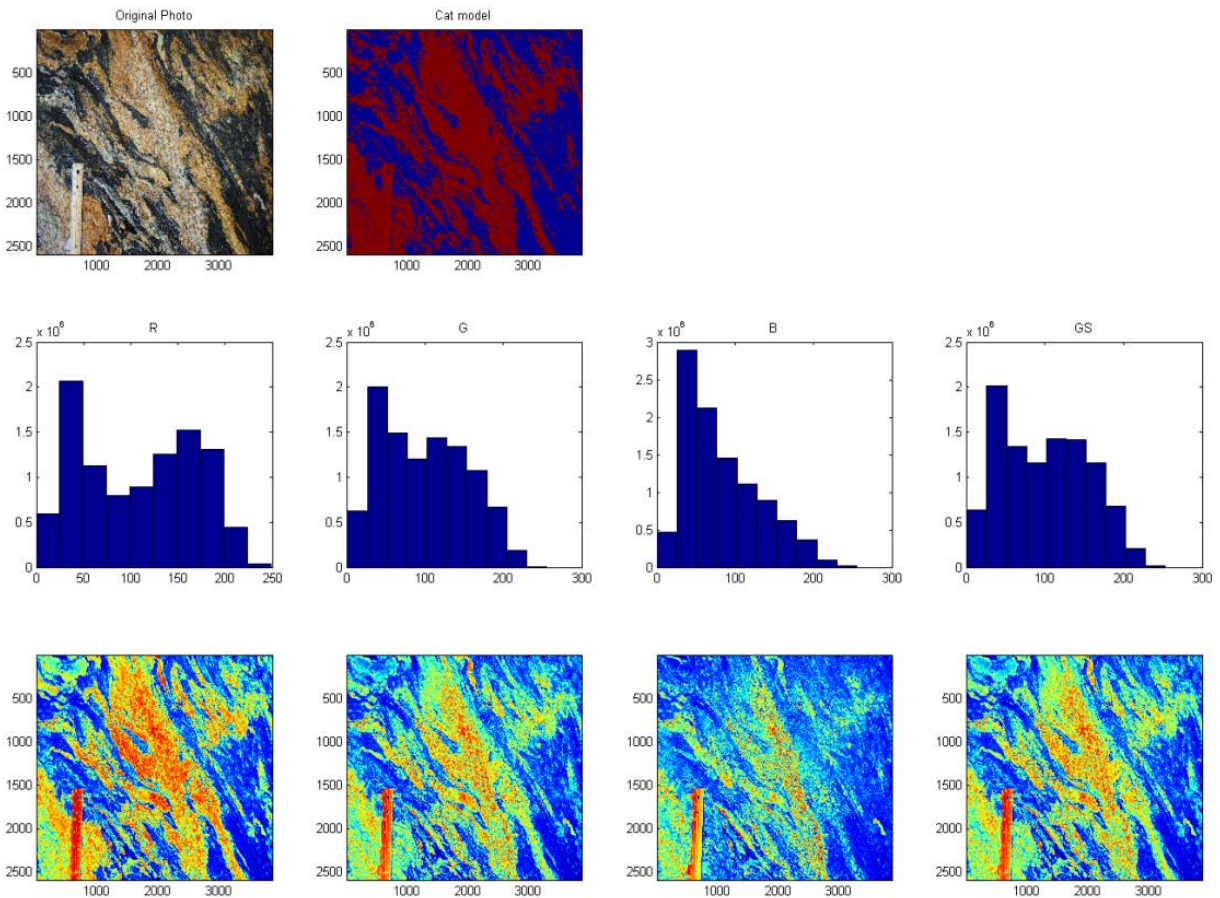


Figure 6. Example subplot of the original photo, categorical model, R, G, B and Gray Scale histogram, and pixelplots of different color values.

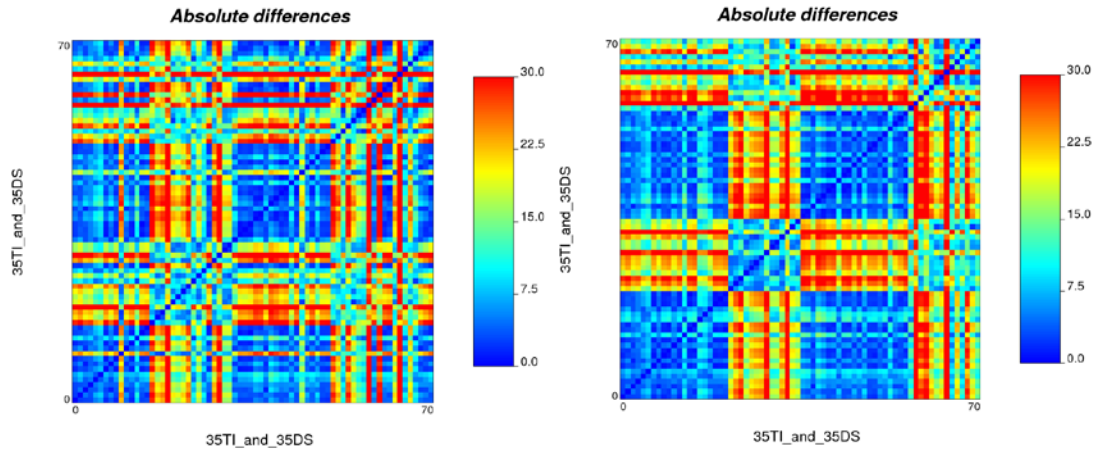


Figure 7. Left: Pixelplot of the absolute differences of all comparisons with original ordering. Right: Pixelplot of the absolute differences of all comparisons after reordering/regrouping. i.e. differences in the lower left quadrant are from comparisons between TIs, differences in the upper right quadrant are from comparisons between drillhole datasets, vice versa.

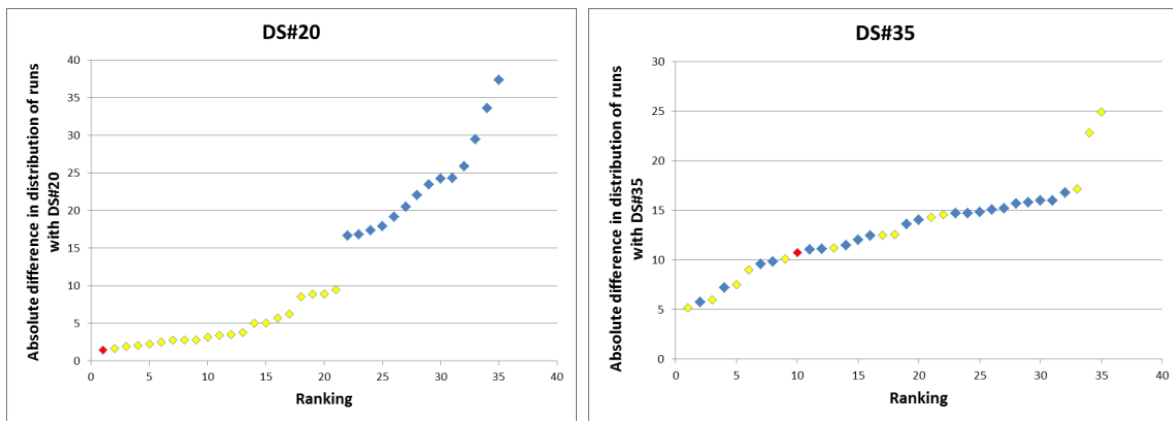


Figure 8. Plots of absolute differences vs. rankings. Note: red dot stands for original TI matching the DS, yellow dots stand for TIs which belong to the same group of the matching TI, blue dots stand for TIs from the other group.

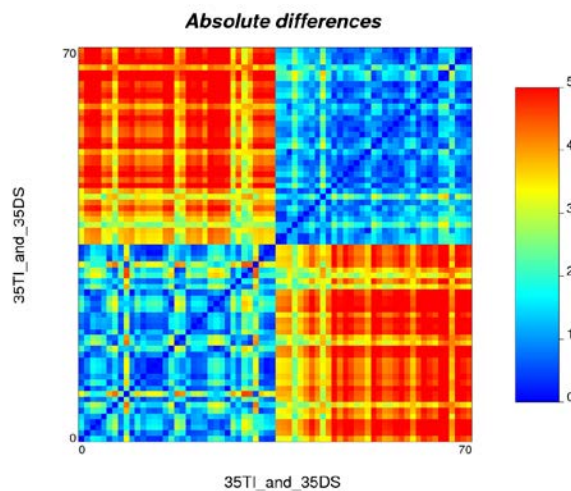


Figure 9. Pixelplot of the absolute differences in MPhistogram of all comparisons with group ordering found by the distribution of runs.

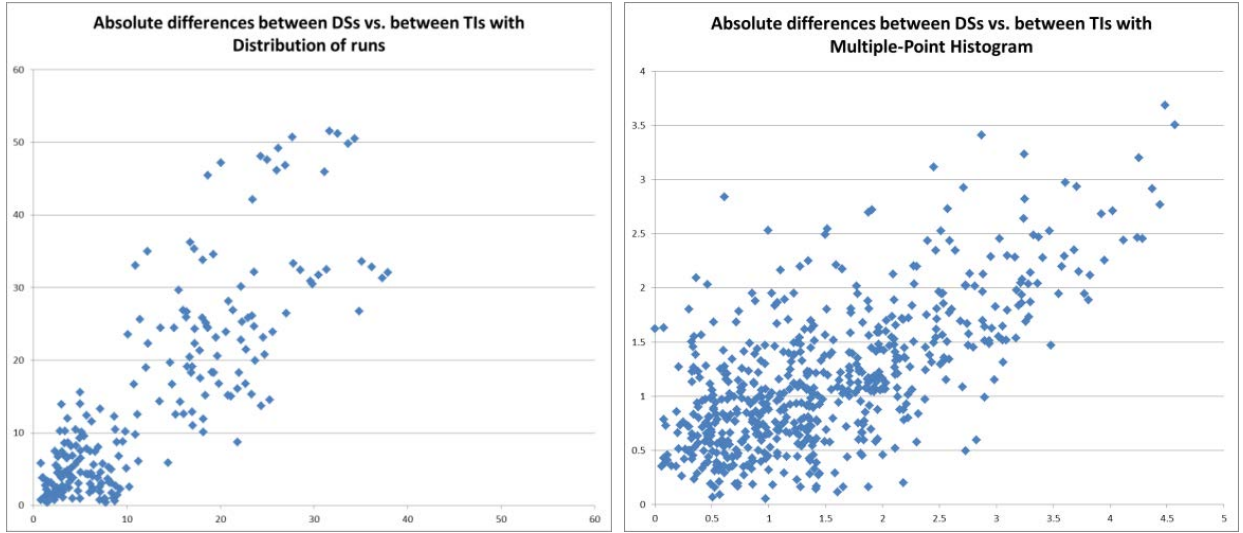


Figure 10. Scatter plot showing the correlations between differences with TI group and DS group, with two different MPS. Correlation factors: 0.832336 with distribution of runs and 0.660469 with MPhitogram.

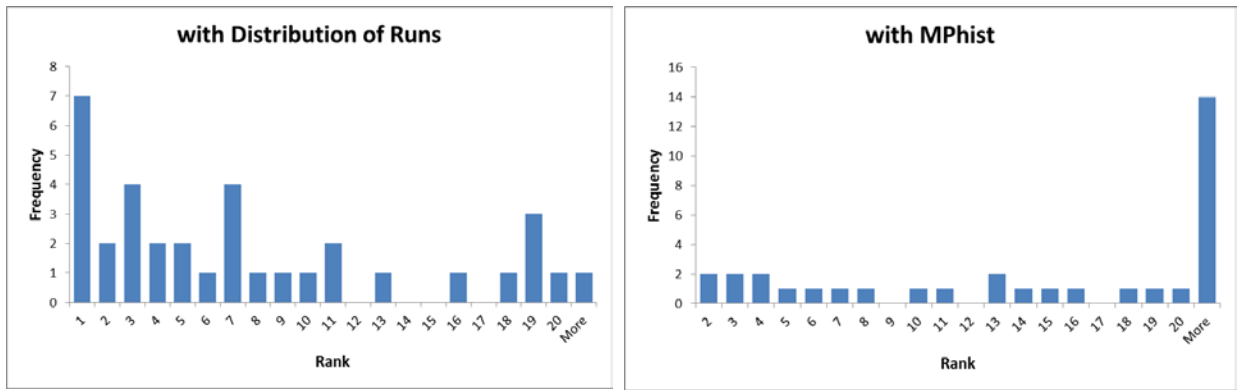


Figure 11. Histograms of the ranking of original TIs to the DSs.