The Concept of Training Image-Based Geostatistics for Spatial Distribution of Hydraulic Conductivity

Fiaz Hussain[1], Ray-Shyan Wu[2]*

[1]*Ph.D. Candidate, Department of Civil Engineering, Water Resources Engineering, National Central University, Taiwan*

[2]*Professor, Department of Civil Engineering, Water Resources Engineering, National Central University, Taiwan*

***Corresponding Author**: engr.fiaz@uaar.edu.pk*

Abstract: Hydraulic conductivity is the key and one of the most uncertain parameters in groundwater modeling. The grid based numerical simulation requires spatial distribution of sampled hydraulic conductivity at un-sampled locations in the study area. This spatial interpolation has been routinely performed using variogram based models (two-point geostatistics methods). These traditional techniques fail to capture the complex geological structures, provides smoothing effects and ignore the higher order moments of subsurface heterogeneities. In this study, multiple-point geostatistics (MPS) technique is used to interpolate hydraulic conductivity data which will be further used in WASH123D numerical groundwater simulation model for regional smart groundwater management. To do this, MPS need 'training images (TIs) as a key input. TI is a conceptual model of subsurface geological heterogeneity which was developed by using the concept of ages, topographic slope as an index criterion and knowledge of geologist. After considerations of full physics of the study area, an example shows the benefits of using MPS compared with two-point geostatistics Kriging for hydraulic conductivity data interpolation in a complex geological formation.

Keywords: Interpolation; Hydraulic Conductivity; Multi-Point Geostatistics; Training Image

Introduction

Understanding the spatial distribution of hydraulic conductivity is a prerequisite for grid based numerical simulation in the field of groundwater resource management. Interpolation methods known as geostatistics [1] are characteristically applied to guesstimate values at un-sampled locations. Traditionally, two-point geostatistics [2] (variogram-based methods) are frequently used to approximate the un-sampled values by evaluating the lower-order statistics of spatial distribution (i.e., covariance). Many researchers employed these methods such as (Kriging, co-Kriging, inverse distance weighting) to estimate un-sampled values in real case studies [3-5]. For example, Theodossiou and Latinopoulos [3] conducted a study in the Anthemountas basin, northern Greece and use the Kriging method for groundwater level interpolation. Kriging and co-Kriging interpolation techniques were applied by Ahmadi and Sedghamiz [4] for groundwater depth mapping. Sun et al. [5] compared three interpolation methods (Kriging, radial basis function and inverse distance weighting) to estimate the groundwater level. Kriging is the optimal technique for groundwater level interpolation in a real case study [5]. In all literature studies, two-point interpolation methods are applied to quantify the spatial correlations only, but for complex formations composed of geologic heterogeneity, these methods interpolation may be inappropriate [6]. These methods cannot reproduce the interconnected, curvilinear geometries characteristic of many heterogeneous complex patterns. These patterns are significant for water resource management.

Multiple-point geostatistics (MPS) [2] approaches offer an opportunity to simulate complex geological patterns. Strebelle [7] was the foremost who successfully implement one of the MPS algorithm known as SNESIM. Hu and Chugunova [8] presented an overview of MPS methods. A book written by Mariethoz and Caers [9] provides a comprehensive introduction to MPS with the help of training images.

Traditional geostatistics talks about hard and soft data and their realization that reflect the patterns and constraints related to these data. While MPS is a method that is qualified within other means of modeling mostly in reservoir modeling, hydraulic conductivity modeling and rainfall history reconstruction [10, 11]. Where traditional variogram modeling is transferred to physical process-based model that is more realistic and very accurate in terms of their physical representation than standard geostatistical techniques. MPS is performed with the help of Training Images (TIs) [12]. A training image (TI), i.e., a conceptual physical model of the geology is the container of patterns as a representation of spatial continuity that has to be simulated, is the fundamental requirement for MPS simulation.

In this study, two-point variogram based estimation (Kriging) and training image base simulation (MPS) is used to produce several stochastic realizations that represent the spatial variation of geologic units in the subsurface. These two methods were compared in this study with the help of example data. In the example, the results show that the MPS better reproduces the patterns of hydraulic conductivity than Kriging, a two-point geostatistics method.

Study Region and Geology

Simply, the training image is the conceptual model of geology consist of hydrology facies of formation represented by patterns. TI can be generated with the help of outcrop data, the concept of ages, topographic slope as an index criterion and knowledge of geologist. In this study, the subsurface cross-section at location CC' of Dashu city (大樹) as shown in Fig.1 of Pingtung county was used as a TI. These subsurface cross-sections were developed by the knowledge of geologist as shown in Fig.2.

Dashu city area consists of four different geological formations. According to geologist, the area consists of Holocene (Alluvium) and Holocene (Terrace Deposits) which is a newly developed formation in the result of earth quicks and erosion process. This is a flat region having mud, sand and gravel mixing. The fraction of this formation is 0.476 in the study region. The second formation is called Pleistocene (Linkou Conglomerates) aggregated with mudstone interbeds, intercalated with sheet and sandstones. The fraction of this formation is 0.419 and consist of the mountainous region. The third formation is known as Pleistocene (Tashe Formation) consist of thick mudstone with sandstone and conglomerates interbeds. This formation fraction is 0.042. The fourth formation is Pliocene (Nanshihlun Sandstone), which consist of thick sandstone, mudstone. The fraction of Pliocene in the region is 0.064. The spatial distribution of formations is shown in Fig.3

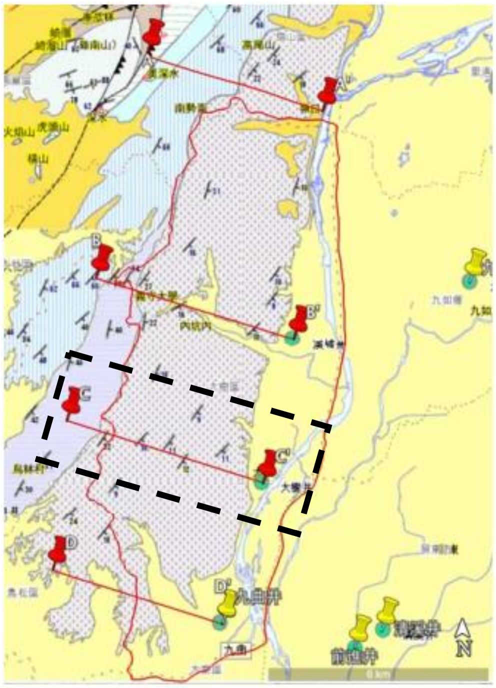


Figure 1: Regional scale geological profile location map of Dashu City

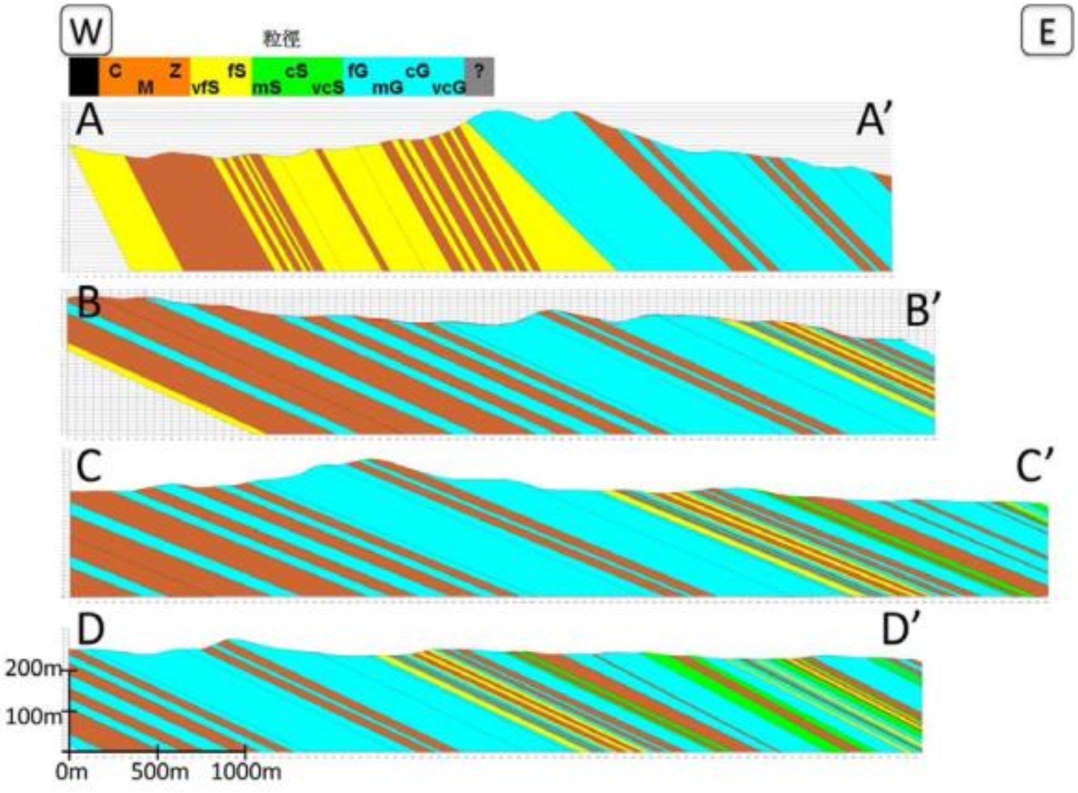


Figure 2: Regional scale subsurface cross-sections of Dashu city

The CC' cross-section was used as a TI which consists of three formations. The boundary of these formations in cross-section was measured with the help of topography and surface boundary of these formations. The cross-sections were developed with the help of borehole data and Líndiànshùn, 1991 study [13]. The distribution of grain size slices consists of clay and mud layer (CMZ) while vfS and fS are the very fine and fine sand layers. The green colour layers (mS+cS+vcS) indicate the medium sand, coarse sand and very coarse sand. The fine, medium and coarse gravel (fG+mG+cG) grains are also in mixed condition as shown in Fig.2

Methodology

Prior Estimation of Hydraulic Conductivity (K) Data

The regional study area consists of a limited amount of data. Only three borehole log data in the Holocene formation help to determine the hydraulic conductivity. Kozeny-Carman Law [14] was used to estimate K (m/s) in Alluvium formation.

$$K = \left(\frac{\rho_w g}{\mu_w}\right) \cdot \left(\frac{d^2}{180}\right) \cdot \left[\frac{\emptyset^3}{(1-\emptyset)^2}\right]$$

Where d is the grain size; ρ_w is the fluid density (1000 kg/m^3); μ_w is the dynamic viscosity taken to be 0.0014 kg/m s [15]; $\emptyset$ is the effective porosity which is estimated with the help of the following formula [16].

$$\emptyset = \emptyset_n(1 - V_{sh})$$

Where, $\emptyset_n$ is the neutron porosity, determine with the help of gamma-ray log (cps) count per second. The relationship between neutron porosity and neutron count rate developed by US geological survey [16] was used to determine the effective porosity. V_{sh} is the volume of shale can be estimated from gamma-ray log [17]

$$V_{sh} = 0.083\left[2^{3.7\left(\frac{GR - GR_{min}}{GR_{max} - GR_{min}}\right)} - 1\right]$$

Where, GR is the gamma ray counts at given depth of layer, GR_{max} and GR_{min} are the maximum and minimum count rate of whole strata.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2021 doi:10.20944/preprints202102.0618.v1

The grain size diameters were obtained from the range of USGS survey for grain size distribution. The range of different diameters according to layers is given below.

CMZ = [0.001, 0.004]; vfS + fS = [0.02, 0.04, 0.08]; mS+cS+vcS = [0.1, 0.2, 0.5, 0.8, 1]; fG+mG+cG = [2, 4, 8] mm.

Based upon the above method, the hydraulic conductivity estimated in alluvium formation is given in Table 1.

Table 1. Estimated Hydraulic Conductivity in Holocene (Alluvium) Formation

Holocene Formation (Alluvium)	
Core Type	K (m/s)
cG+mG+fG	1.05×10^{-3}
vcS+cS+mS	6.60×10^{-4}
fS+vfS	2.95×10^{-5}
C+M+Z	1.26×10^{-6}

The hydraulic conductivity of other formations was considered by the hypothesis of alluvium to the non-alluvium formation. It was supposed because of the non-availability of borehole data. The hydraulic conductivity of Pleistocene (Linkou Formation) was assumed one order reduction in the magnitude of Holocene (Alluvium) formation and for Pliocene (Nanashihlun) formation two order reduction in the magnitude of Holocene. For Pleistocene (Tashe) formation, the hydraulic conductivity was assumed half order magnitude reduction from Pleistocene (Linkou) formation. The hydraulic conductivity of non-alluvium formations is given in Table 2.

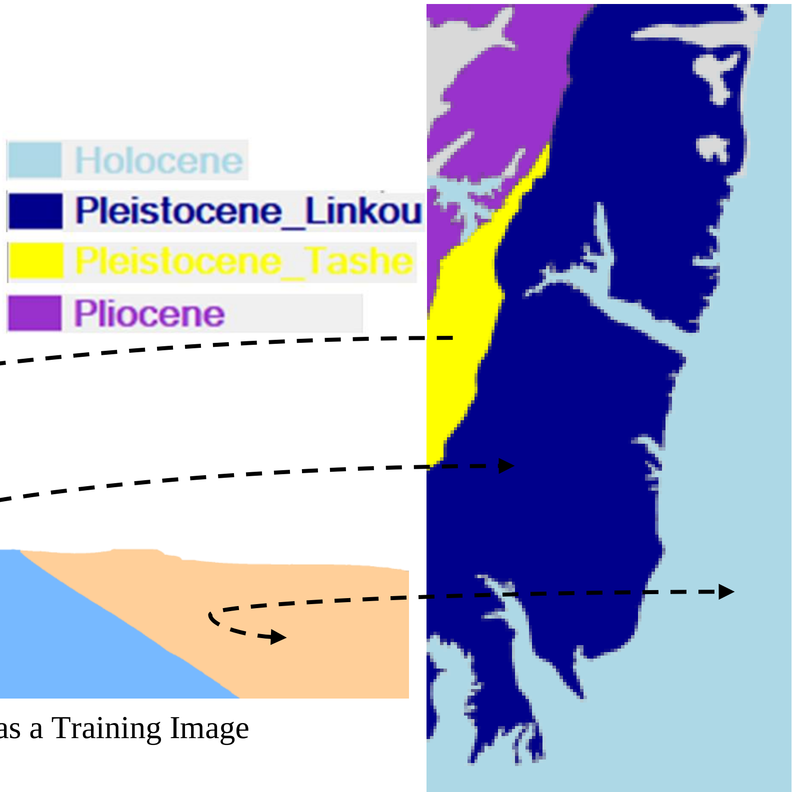
Table 2. Hydraulic Conductivity of Non-Alluvium Formations

Pleistocene (Linkou)		**Pleistocene (Tashe)**	**Pliocene (Nanshihlun)**
Core Type	K (m/s)	K (m/s)	K (m/s)

cG+mG+fG	1.05×10^{-4}	5.25×10^{-5}	1.05×10^{-5}
vcS+cS+mS	6.60×10^{-5}	3.30×10^{-5}	6.60×10^{-6}
fS+vfS	2.95×10^{-6}	1.47×10^{-6}	2.95×10^{-7}
C+M+Z	1.26×10^{-7}	6.30×10^{-8}	1.26×10^{-8}

Training Image and Integration of Hydraulic Conductivity Data

The CC' cross-section was scanned and used as a training image as shown in Fig.4. The boundary of the cross-section is marked with the help of Fig.3. The training image has three categories, which were indicated as 0, 1 and 2 values. The histogram of these categories indicated the correct target marginal distribution, which was found as 0.30, 0.54, and 0.16, respectively as shown in Fig.5



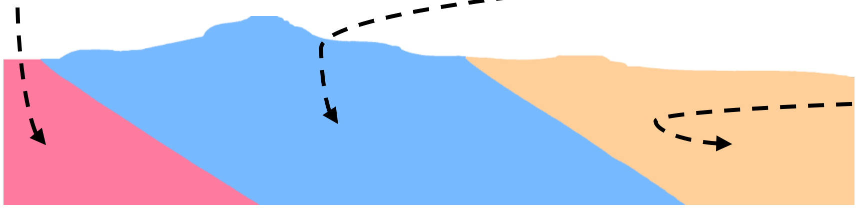


Figure 4. CC' Cross-section used as a Training Image

Figure 3. Surface geology map of study area

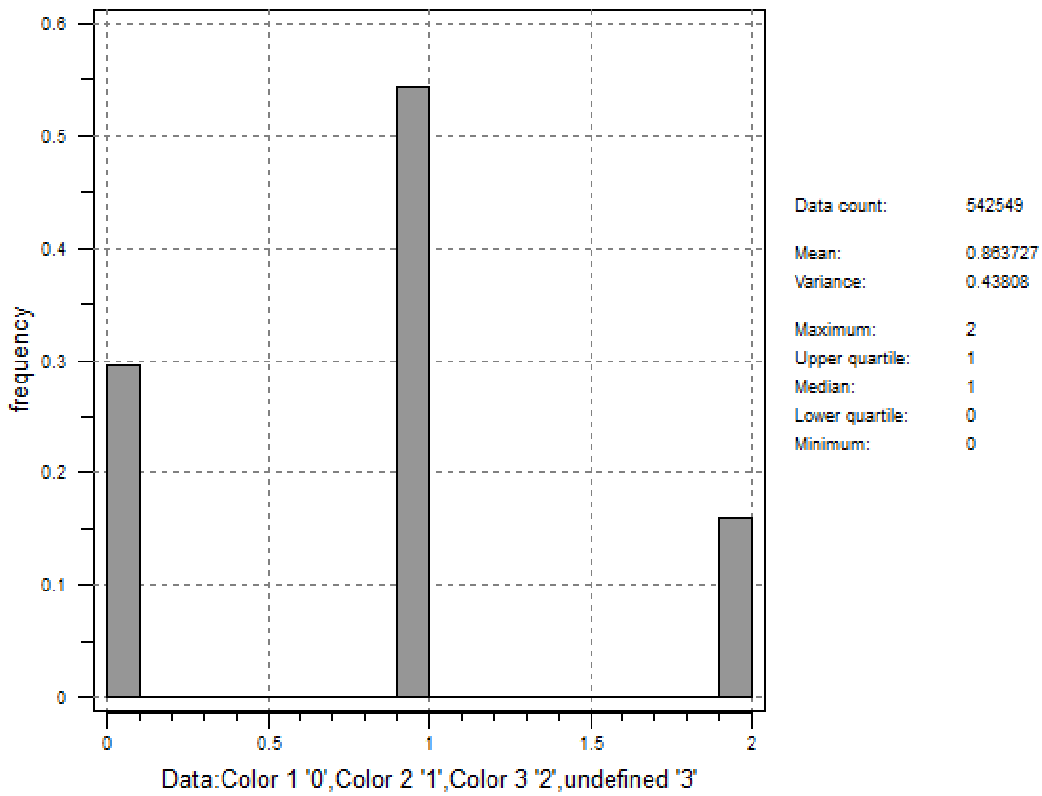


Figure 5. Target Marginal distribution of categories of Training Image

The CC' cross-section was also categorized into different classes as shown in Fig.6 for the incorporation of prior estimated hydraulic conductivity data. For this purpose, this 2D cross-section (XZ-direction) were converted into a grid with size: 2130 x 1 x 310 with a unit cell size in each direction. The blue colour is the background and used as a No Data (-9966699). This grid data is used as hard data for the pattern determination in the kriging process and further for conditioning in the MPS method.

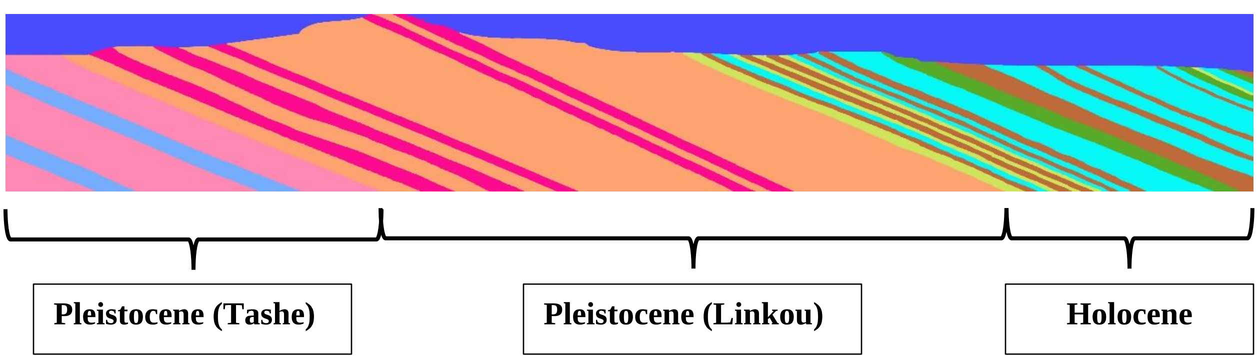


Figure 6. Hydraulic Conductivity Grid Data of Cross-section

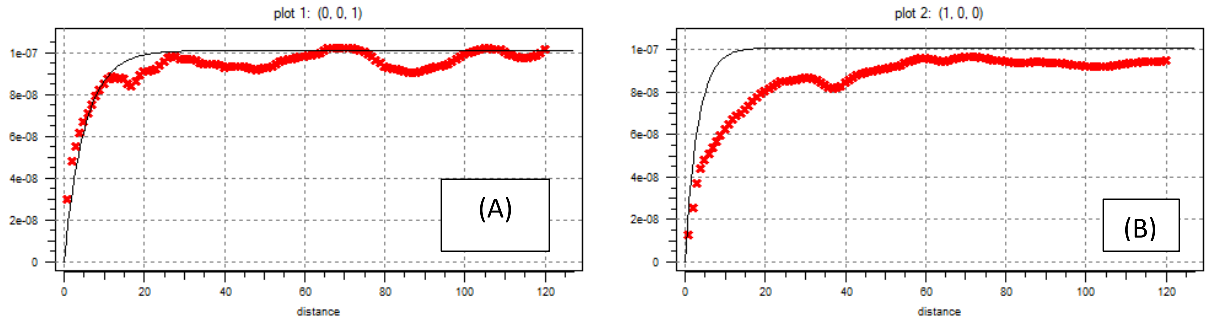
Estimation and Simulation using SGeMS Software

The Stanford Geostatistical Modeling Software (SGeMS) is a general-purpose, user-friendly, state-of-the-art geostatistical software package. In this study, SGeMS is used to estimate and simulate hydraulic conductivity data at un-sampled locations using Kriging and MPS, respectively. Currently, there is only one practical software that is perfectly available to do MPS is SGeMS. The software code is in the public domain, downloadable from http://sgems.sourceforge.net/

Results and Discussion

The variogram, a 2-point statistics

The estimate algorithm of Kriging is performed in SGeMS using variogram 2-point geostatistical modeling. The variogram or its equivalent (covariance) is the main tool for kriging. The experimental variogram was developed in two directions. The x-direction is the minor direction while the z-direction is the main direction. Half of the number of cells of the main direction was considered as No. of lag. For vertical variogram, we use dipping of 90 degrees of main data. The model and experimental variogram are shown in Fig.7



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2021 doi:10.20944/preprints202102.0618.v1

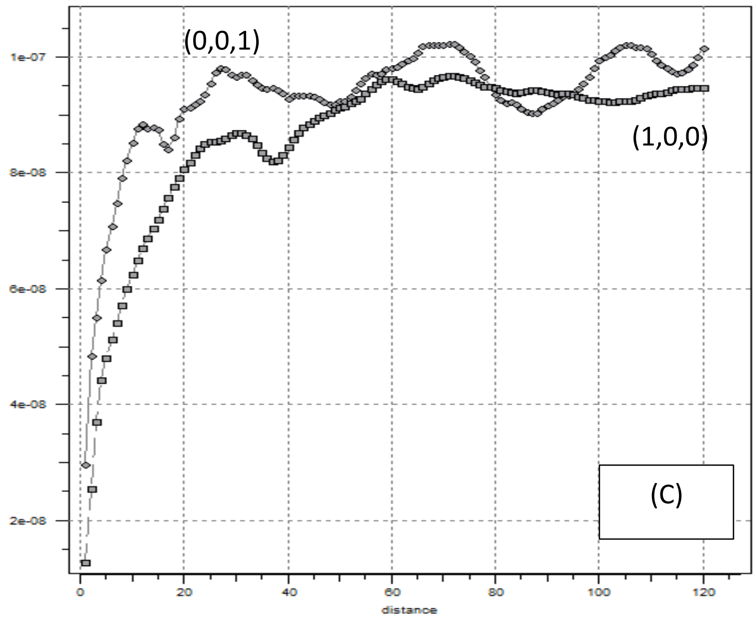


Figure 7. Experimental variogram (A) Z-direction (B) X-direction and (C) Both Direction

Historically, kriging remains a major integration tool in estimation and simulation algorithms. It interpolates the data-to-unknown correlation to data-to-data correlation through a non-diagonal kriging matrix. In this work, the Ordinary Kriging (OK) algorithm is used. In OK, the predictive value of the random function is locally re-estimated from local data, while the covariance model is kept stationary. The variogram in Fig.7 looks nice and smooth with range, sill and nugget effect. The analysis indicated the geometric anisotropy because there is a change in range with direction. The variogram is subject to sampling variance indicated no over modeling effect as shown in Fig.7. The first few lags are very important for kriging results. Here the lags look linear and smooth, so spherical variogram type is used.

The OK results are shown in Fig.8. For performing the OK, the search ellipsoid is very import for major continuity. Here used single search neighbourhood in search ellipsoid and dip by 90 degrees to get maximum vertical data. To avoid the error, the conditional dataset was increased in data of search neighbourhood.

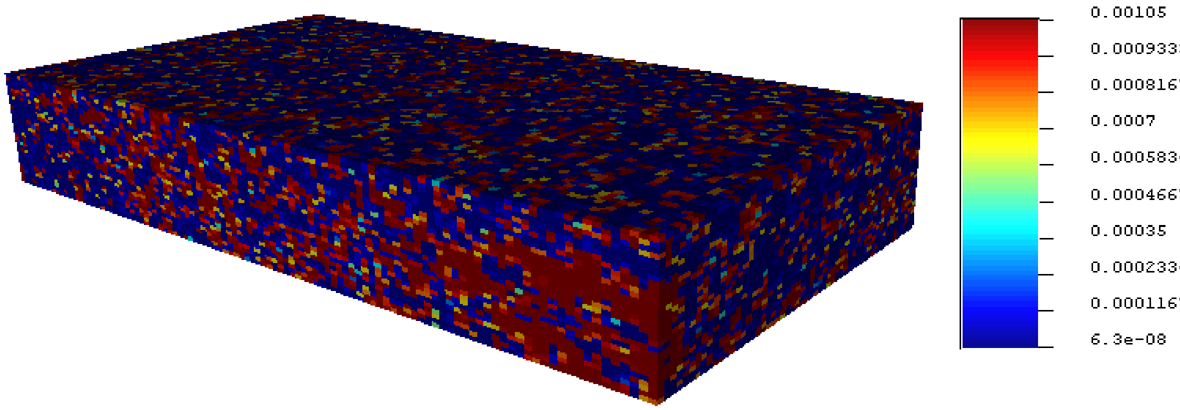


Figure 8. The Ordinary Kriging (OK) estimation results

The OK is performed using hydraulic conductivity data of cross-section CC' on a grid system having a size (107 x 55 x 33), cell size (20, 20,10). The interpolation results obtained are smooth indicated the use of two-point spatial correlation within the data and could not simulate the complex patterns as shown in the training image. Hence 2-point geostatistics can only yield a variogram model and failing to reproduce definite shape and patterns.

Multi-point Simulation using SNESIM Algorithm

The MPS concept became practical with the SNESIM (Single Normal Equation Simulation) implementation of Strebelle (2000) [11]. In SNESIM, the Ti is scanned only once; all conditional proportions available in that Ti for a given search template size are stored in a search tree data structure, from which they can be efficiently retrieved. This algorithm contains two key parts, the construction of the search tree which stores all training proportions, and the simulation part itself. In this study, a SNESIM is performed as a 3D simulation conditioning to hydraulic conductivity data. The large 2D training image of cross-section CC' is used. The training image dimension (2130 × 1 × 310), and facies proportions 0.30, 0.54, and 0.16 for Holocene, Pleistocene Linkou Formation and Pleistocene Tashe Formation, respectively. The hydraulic conductivity point data of the CC' cross-section provide hard conditioning data at each location. The simulated field is of size 107×28×40 with cell size 20, 20, 5. For SNESIM simulation, 60 conditioning data nodes are retained in the search template. The ranges of search ellipsoid in three major axes are 100, 100 and 5, respectively. The angles for azimuth and rake are zero while the dip is 90 degrees. No use of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2021 doi:10.20944/preprints202102.0618.v1

affinity or rotation. 4 multiple grids are selected with isotropic template expansion. For data conditioning, 4 additional nodes were used in the subgrid concept. The servosystem value is 0.9. One SNESIM realization conditioning hard data is given in Fig. 9.

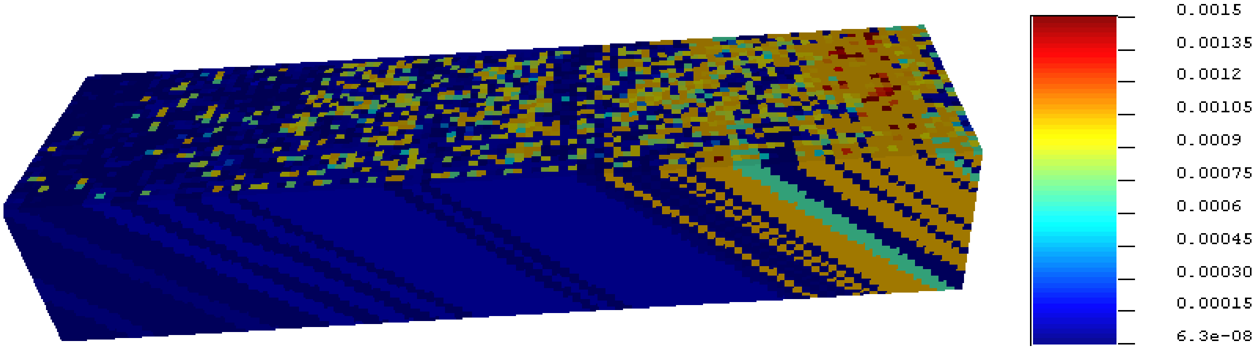


Figure 9. One SNESIM realization of hydraulic conductivity data

The simulated map using MPS exhibits similar patterns to the true map. The computational cost of the MPS method for the interpolation is much higher than that of Kriging.

Conclusion and Recommendation

The question “which model, 2-point or MPS?” is very important for the selection of modeling technique. The simple criterion is “The better model is that which delivers the “deemed” better result”. In this study, SNESIM is mimicking the true map while OK is providing smooth results. Hence, the TI patterns yield a “more” accurate result than the simplified structures by variogram-based statistics. For future work, MPS will be used for geostatistical modeling of Dashu City. The surface geology map will be used as a 2D horizontal training image and 1D vertical variogram from four cross-sections for characterizing the geologic heterogeneity in three dimensions which will be further used in WASH123D numerical groundwater simulation model.

References

[1] Varouchakis, E.A. *Geostatistics: Mathematical and Statistical Basis*. In: Spatiotemporal Analysis of Extreme Hydrological Events, Elsevier, 2019, pp. 1-38, https://doi.org/10.1016/B978-0-12-811689-0.00001-X.

[2] Tahmasebi, P. *Multiple Point Statistics: A Review*. In: Daya Sagar B., Cheng Q., Agterberg F. (eds) Handbook of Mathematical Geosciences. Springer, Cham. 2018, pp. 613-643, https://doi.org/10.1007/978-3-319-78999-6_30

[3] Theodossiou, N.; Latinopoulos, P. Evaluation and optimisation of groundwater observation networks using the kriging methodology. Environ. Model. Softw. 2006, 21, 991–1000.

[4] Ahmadi, S.H.; Sedghamiz, A. Application and evaluation of kriging and cokriging methods on groundwater depth mapping. Environ. Monit. Assess. 2008, 138, 357–368.

[5] Sun, Y.; Kang, S.; Li, F.; Zhang, L. Comparison of interpolation methods for depth to groundwater and its temporal and spatial variations in the minqin oasis of northwest china. Environ. Model. Softw. 2009, 24, 1163–1170.

[6] Li, L., Huang, G. Groundwater level mapping using Multi-point Geostatistics. Water 2016, 8, 400; doi:10.3390/w8090400

[7] Strebelle, S. Conditional simulation of complex geological structures using multiple-point statistics. Math. Geol. 2002, 34, 1–21.

[8] Hu, L.; Chugunova, T. Multiple-point geostatistics for modeling subsurface heterogeneity: A comprehensive review. Water Resour. Res. 2008, 44, doi:10.1029/2008WR006993.

[9] Mariethoz, G., Caers, J. Multiple-point Geostatistics: Stochastic Modeling with Training Images, Wiley-Blackwell, 2014, pp.376.

[10] Huysmans, M., Dassargues, A. Application of multiple-point geostatistics on modelling groundwater flow and transport in a cross-bedded aquifer (Belgium). Hydrogeol J 17, 1901 (2009). https://doi.org/10.1007/s10040-009-0495-2

[11] Marini, M., Felletti, F., Beretta, G.P., Terrenghi, J. Three Geostatistical Methods for Hydrofacies Simulation Ranked Using a Large Borehole Lithology Dataset from the Venice Hinterland (NE Italy). Water 2018, 10, 844; doi:10.3390/w10070844

[12] Mohan Srivastava, R. The Origins of the Multiple-Point Statistics (MPS) Algorithm. In: Daya Sagar B., Cheng Q., Agterberg F. (eds) Handbook of Mathematical Geosciences. Springer, Cham. 2018, pp. 655-672; https://doi.org/10.1007/978-3-319-78999-6_32

[13] Lin Dianshun, 1991, Sedimentary facies and sedimentary environment evolution of the Upper-Phoenix in the Laoshan belt in southwestern Taiwan, and a master's degree in the Institute of Geology, National Taiwan University. (in Chinese)

[14] Kruczek, B. Carman–Kozeny Equation. In: E. Droli, L. Giorno (eds.), Encyclopedia of Membranes, Springer-Verlag Berlin Heidelberg 2014, DOI 10.1007/978-3-642-40872-4_1995-1

[15] Fetter, C.W., 1994. Applied Hydrogeology, third ed. Prentice-Hall, Inc., New Jersey, p. 600

[16] Hudson J.D. 1996, Use of geophysical logs to estimate the quality of ground water and the permeability of aquifers. U.S. Geological Survey, Water-Resources Investigations Report 95-4300.

[17] Larionov VV (1969) Radiometry of boreholes (in Russian). Nedra, Moscow

[18] Strebelle, S. 2000, Sequential simulation drawing structures from training images, Ph.D. thesis, Stanford University, Stanford, CA.