

IECMS  
2020

# The 2nd International Electronic Conference on Mineral Science

16–30 NOVEMBER 2020 | ONLINE



## Univariate Geostatistical Outlier Detection Methods Based on Variogram Pairs, Case Study: Sarigunay Gold Deposit, Iran

Simin Saadati<sup>1</sup>, Mohammad Fahimi Nia<sup>2</sup>, and Omid Asghari<sup>3,\*</sup>

<sup>1</sup> BSc Graduate, Simulation and Data Processing Lab, School of Mining Engineering, Faculty of Engineering, University of Tehran, Iran

<sup>2</sup> PhD Candidate, Simulation and Data Processing Lab, School of Mining Engineering, Faculty of Engineering, University of Tehran, Iran

<sup>3</sup> Associate Professor, Simulation and Data Processing Lab, School of Mining Engineering, Faculty of Engineering, University of Tehran, Iran

\* Corresponding author: o.asghari@ut.ac.ir



**Abstract:** Statistically, outliers are the data remarkably dissimilar to the whole dataset. These existing outliers may give rise to misinterpretations in statistical and geostatistical analyses. To detect outliers two methods of (1) boxplot as a representative of statistical methods and (2) a combination of Mahalanobis Distance (MD) and network graph as a representative of geostatistical methods are applied. In this case, lags of 20 meters are applied to the h-scatter plot. Then, Mahalanobis distance and 97.5% confidence interval, taken from chi-square distribution, are applied to the H-scatter plot to detect pairs of outliers. In order to consider geospatial relation of each pair, a network graph is designed which counts the number of edges for each node. The number of edges demonstrates the outliers and their neighbouring nodes which the outlier detection is based on. The mentioned process, applied to the oxide zone of Sarigunay epithermal gold deposit in Iran, results in 286 data points detected as outliers through an 11945 sampled dataset in which the ratio of outliers to raw data is 2.39%. The boxplot drawn for the raw data indicates the cut-off assay of 10 ppm Au. Substantially, combination of statistical and geostatistical outlier detection methods leads to robust variograms and more precise estimation.

**Keywords:** Outlier Detection, Variogram Pairs, Network Graph, Mahalanobis Distance, Sarigunay

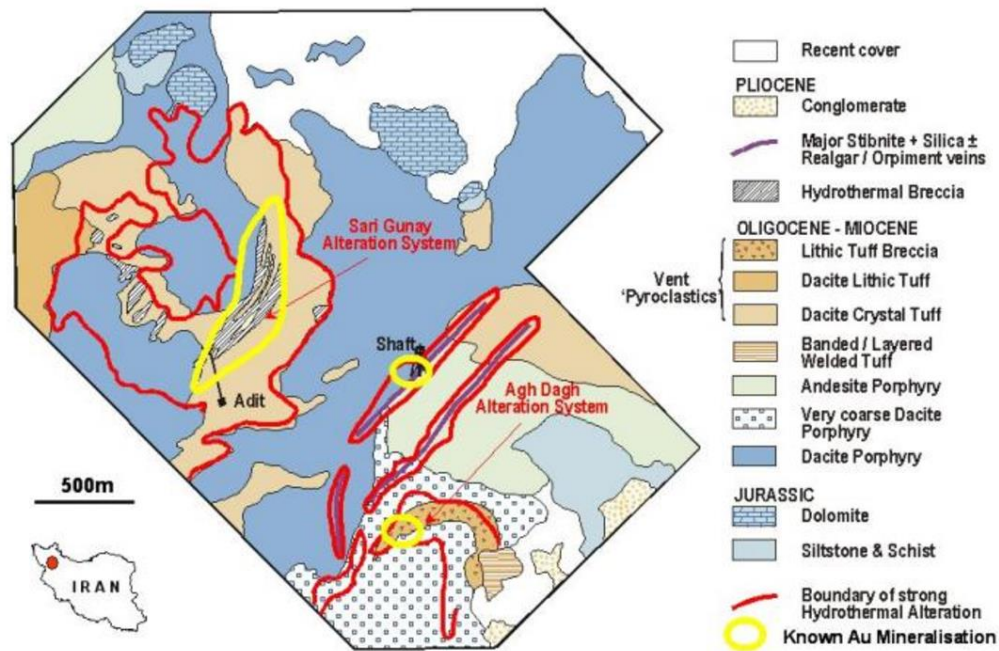
# Case Study

## *Sarigunay gold deposit*

Epithermal gold deposit – Oxide zone

11945 observations

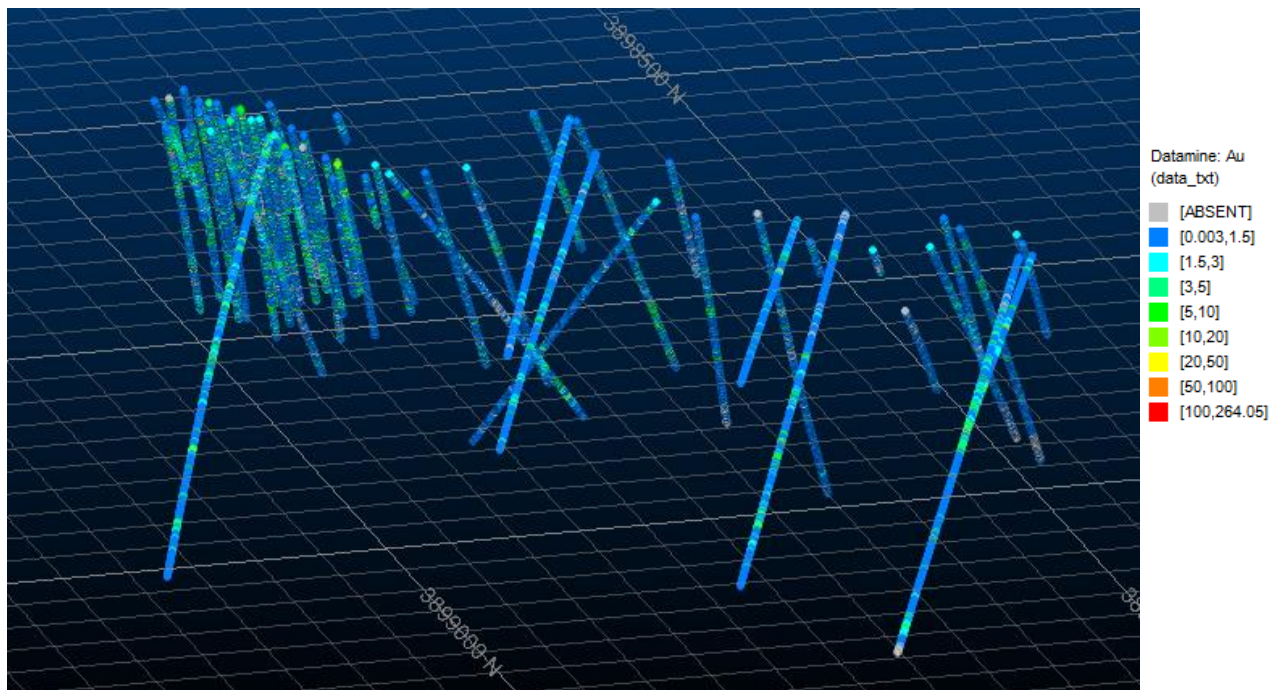
Au assays



# Case Study

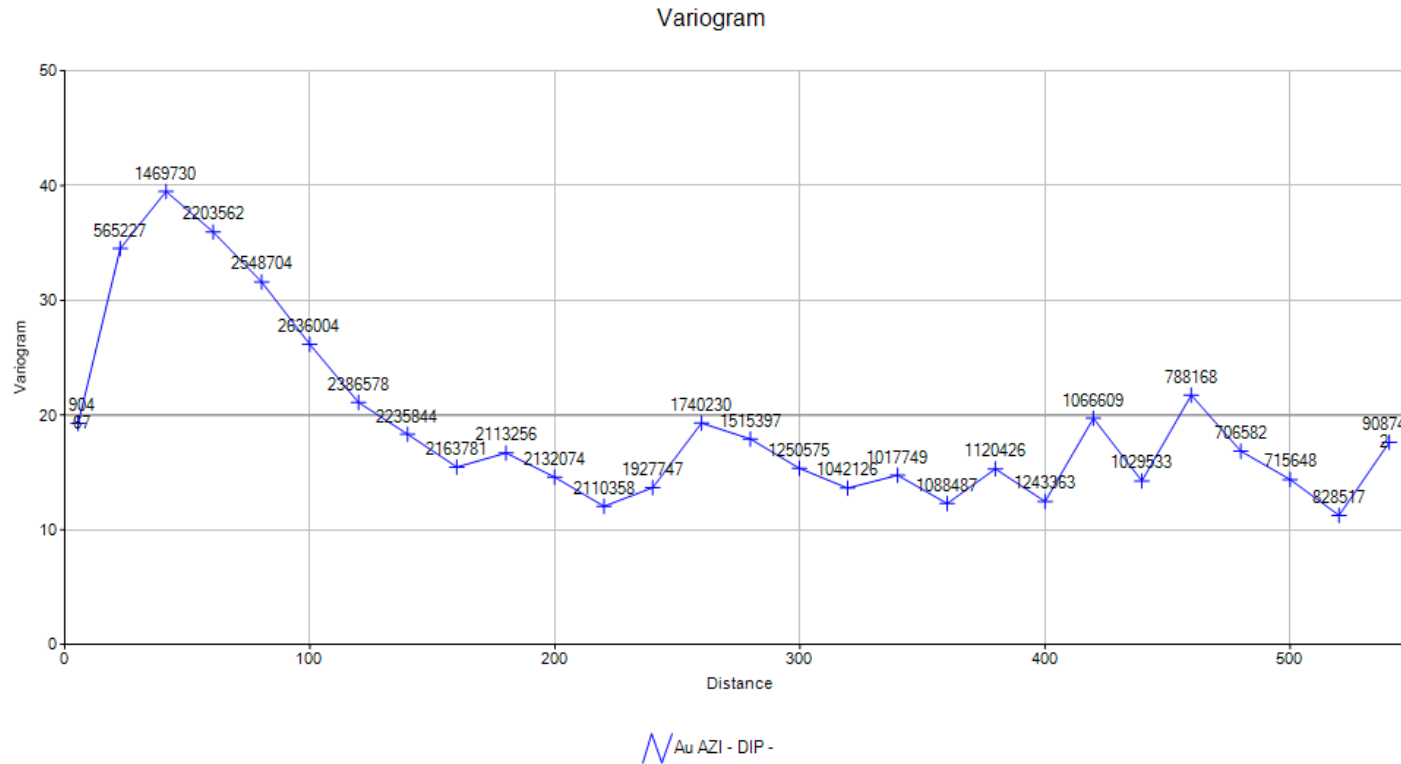
## Descriptive Statistics of The Dataset

Variable	Count	Minimum	Maximum	Mean	Std. Dev.	Variance	Skewness	Kurtosis
Au	25269	0.002	263.95	1.446	3.661	13.401	21.1	1146.804



NW-SE Section of Boreholes

# Variogram

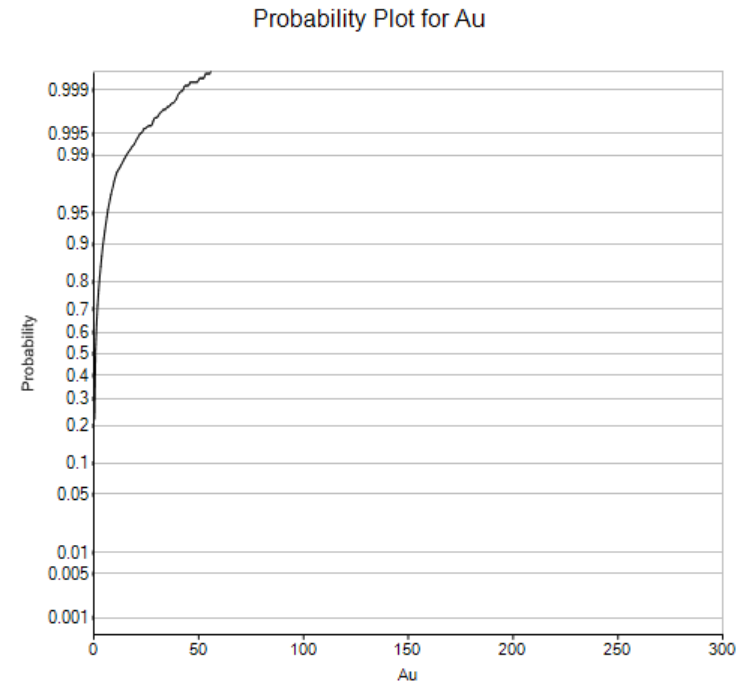
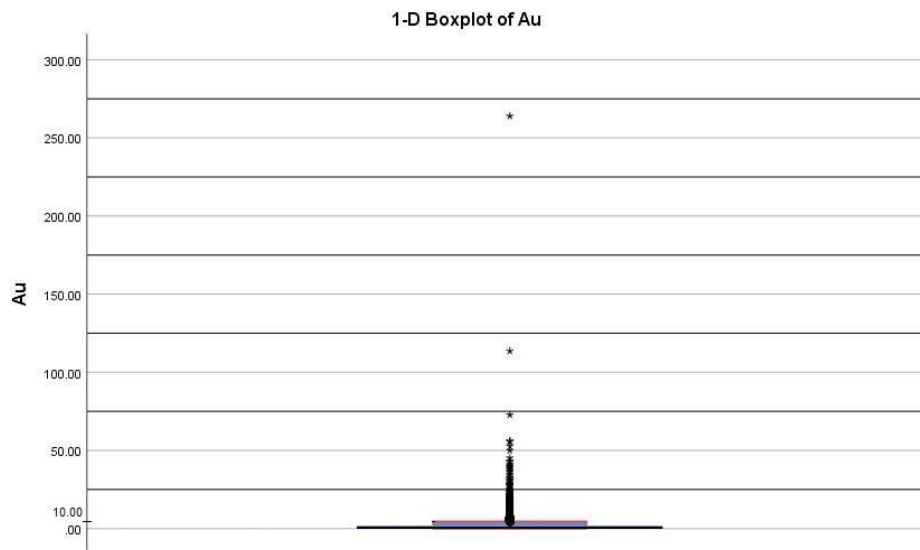


Omni-dimensional Variogram of Gold Oxidation Zone

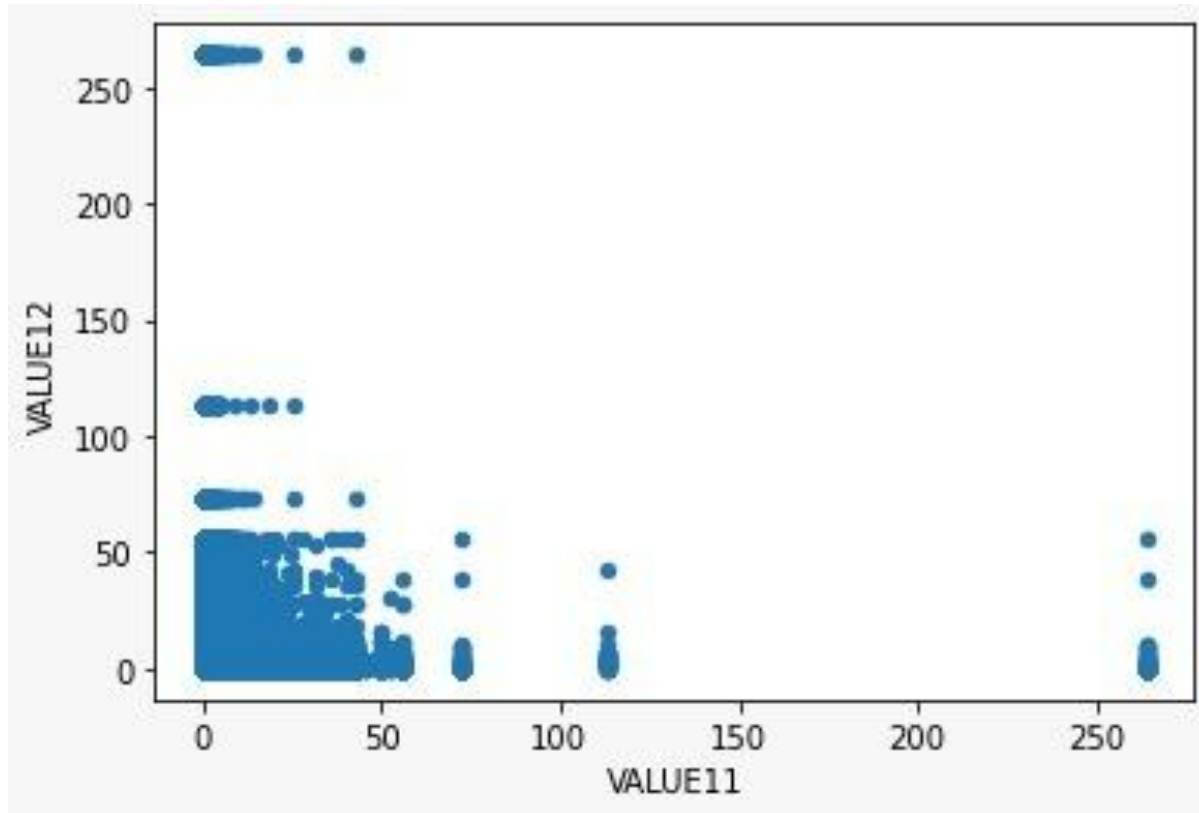
IECMS  
2020

# Statistical Detection of Outliers

- Box-plot and pp-plot applied to the data
- Cut-off 10 ppm deduced for outliers threshold



# H-scatter Plot



H-scatter Plot For Lags of 20 Meters

# Mahalanobis Distance & Network Graph

- Allocate a distance to each observation
- Robust distance → S-estimator
- Statistical mean of the data
- Computing the covariance matrix

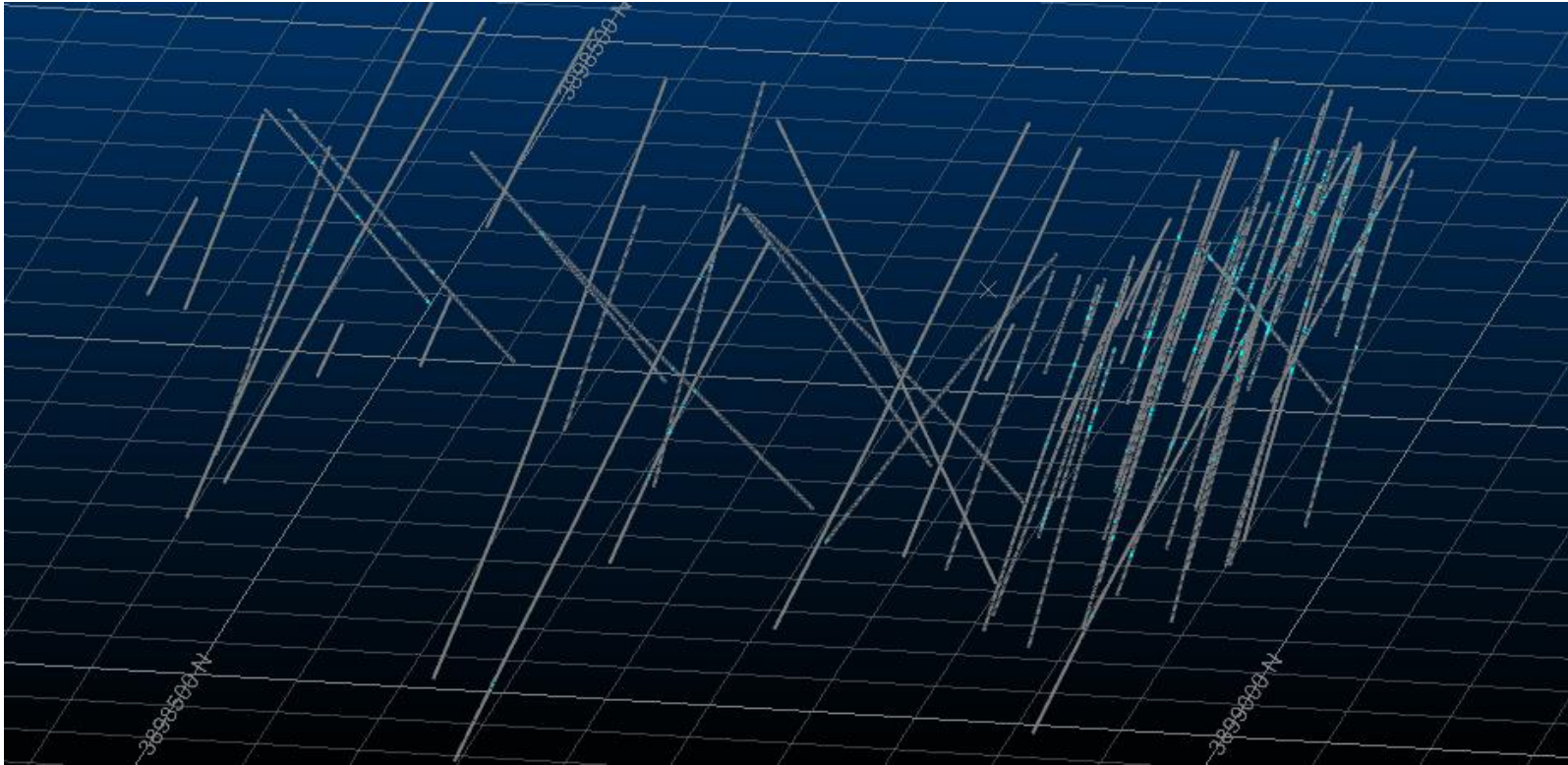
$$D_M(\vec{x}) = \sqrt{(\vec{x} - \vec{\mu})^T S^{-1} (\vec{x} - \vec{\mu})}$$

Mahalanobis distance has been applied to the H-scatter plot and critical value (by 95% confidence interval) from the chi-square distribution contributed in calculation of the outlier pairs.

To 1107 pairs of detected outliers, the network graph was applied and 286 data points with edges (i.e. connections between data points in less than 20 meter distance) more than 3 were selected.



# Result



Representation of Geostatistically Detected Putliers (blue color)

# Conclusions

Detection of outliers using statistical methods is solely based on the values. However, spatial methods of outlier detection consider the distance between data points only. The combination of statistical and spatial methods, known as geostatistical methods, result in more practical outlier detection.

Using Mahalanobis distance as a robust method for statistical outlier detection with network graph as a basic spatial method designate both statistical and spatial aspects of the outlier detection. The combination of these two methods is known as a geostatistical outlier detection.

Substantially, 286 data points detected as outliers through an 11945 sampled dataset in which the ratio of outliers to raw data is 2.39%.