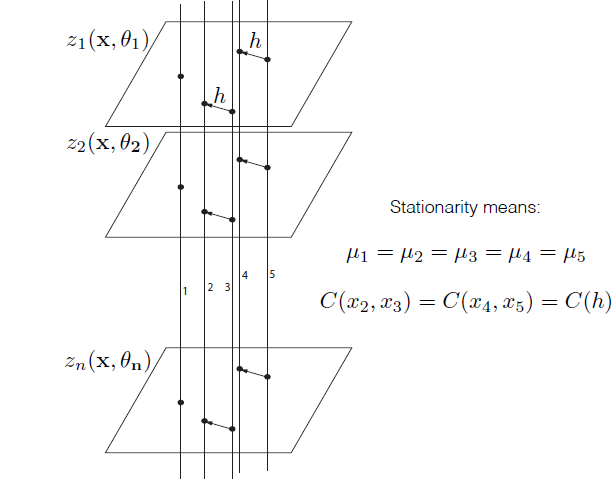
# **Spatial relationships: Estimation and Modelling**

## Random Function Model

### Requirement of stationarity

A random function model is *loosely* stationary when the mean is constant and the covariance is translation invariant. (For full stationarity all the *central* moments has to be tested?)

In geostatistics, we mostly deal with Gaussian distribution (any other distribution is (normal score) transformed to Gaussian/Normal distribution). Gaussian distribution has only two central moments and when both are constant; we can say the Gaussian distribution is fully stationary.

Mathematically, the first order stationarity can be written as:

The second order stationarity can be defined as:

any function of two random variables distance apart is independent of the location and only dependent on the distance and direction (if anisotropic).

Covariance is the commonly used function relating the two variables.

## Spatial relationship

Spatial relationships are used to describe how neighbouring values are related. Three main relationships are: covariance, correlation coefficient & variogram. The dataset can either belong to one variable X spatially distributed or multiple variable X, Y spatially distributed.

### Covariance

In practice (on available sample data):

Where assumed to be 1st and 2nd order stationary

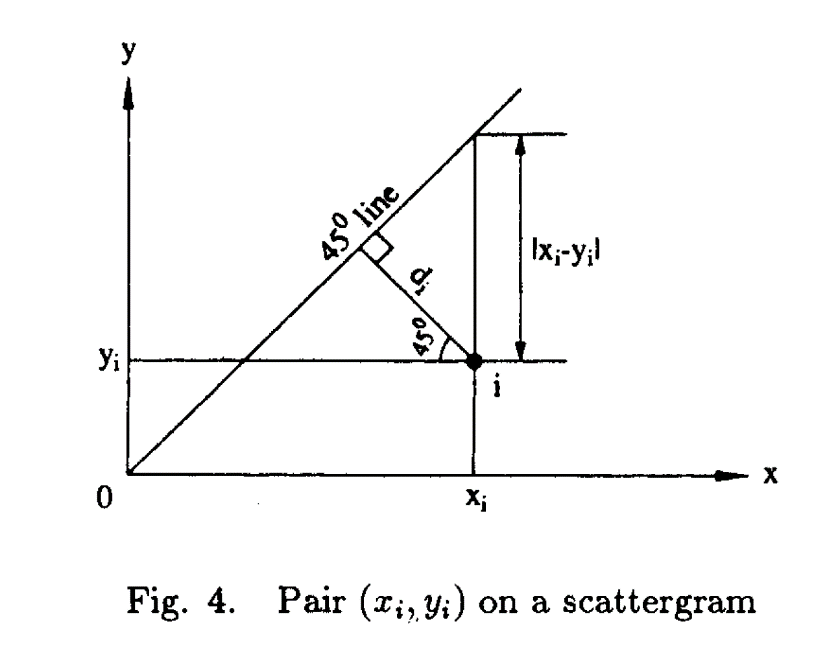
### Correlation Coefficient

For entire populations:

Where

### Variogram

#### What is variogram?

Semivariance (*moment of inertia*) is calculated, as shown in the scatterplot alongside:

1. Plot the variables in the scatterplot
2. Calculate the vertical distance from the x=y line
3. Now calculating the average of

Thus, the moment of inertia of the scattergram around e.g. the 45° line would be a characteristic of lack of dependence. This moment of inertia is called the "semi-variogram" of the set of data pairs and the variogram, is none other than the average of squared difference between the two components of each pair.

When the semivariance of each lag distance is plotted we get the variogram.

Note: Another way of looking at variogram. (From Kelkar, Perez)

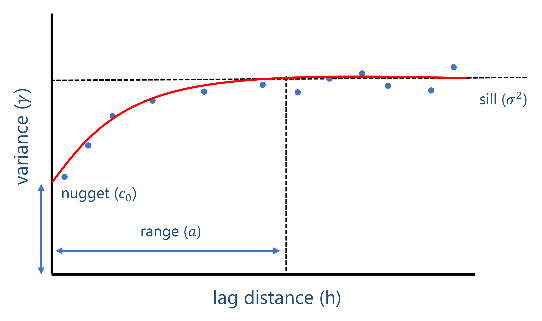
It is *half* of the *variance* of the difference of the value of the variable at two locations located distance apart.

#### Relationship between Variograms and Covariance

=

OR

Or variogram (**semivariance** between 2 points) = **sample variance**  – **covariance** between the 2 points. (at range, the semivariance reaches the max. sample variance)

Steps to get a variogram:

* 1. For each lag distance , the is calculated or just 1 point is plotted. (see diagram alongside)
  2. With changing lag distance lag distance , all the points are plotted to get the variogram.

#### Relationship between variogram and covariance (for non-ergodic systems)

For practical purposes (from the data set), the estimated variogram is:

**&**

**&**

Hence, by substituting the above variables in the equation above: (no stationarity principle applied)

The above equation can only turn into when the following stationarity holds true:



The above relations hold true only when we have sufficient pairs at each lag distance.

If we don’t have sufficient pairs, sample mean and variance can be different from lag mean and variance. In that case, we use the **non-ergodic variogram**.

Non-ergodic correlogram:

It accounts for the variation in the sample and lag variances.

## Estimation of Variogram

### Lack of Sufficient Pairs

#### Number of Pairs

For a fixed number of data points, the pairs keep decreasing with the increasing lag distance and so does the confidence in our data.

Thumb Rule: The variogram calculations should be restricted to *½ of the max distance* between any two data points within the ‘region of interest’ where the stationarity is assumed.

#### Lag Tolerance

Another alternative to ensure sufficient number of pairs for a given lag distance, is to describe a tolerance with respect to distance as well as direction/azimuth. (Because in a spatial distribution samples are not distributed at exactly)

Any point in the space can be taken into the calculation of the semivariance of 2 consecutive lag distances. E.g. a point come in the lag 32 and lag 42 calculations.

Thumb rule: Usually one should start variogram calculation with a small tolerance. If too many fluctuations, we will keep adjusting the tolerance until we get an interpretable structure. E.g., ratio of the range in the x-direction to y-direction is 5, which reduces to 2.4 once the directional tolerance is increased to 45 degrees.

### Instability of variogram

Since the estimated variogram is the arithmetic average of the squared differences of the two variables, any big differences when squared can have a significant impact on the semivariance value.

Two ways it can be reduced:

* Increasing the number of pairs at a given lag distance
* Remove certain extreme pairs (with very large differences). Cross plot the two variables (for a certain lag distance

### Influence of Outliers

#### Log Transform

Using the logarithmic values of the samples

#### Power Transform

Using the power values of the samples:

#### Rank Transform

All the values are set in ascending order and assigned the Rank, which are then used as values

#### Indicator Transform

The indicator transform allows us to transform the continuous variable into a discrete variable.

##### Indicator Variable

Indicator variable

So indicators can take either 0 or 1. By setting different thresholds, we can get multiple indicator values.

###### Expected Value

Cumulative distribution function

###### Variance

Since can only take values of 0 or 1, .

So,

###### Non-centred covariance

One of the unique properties of the indicator variable is its ability to define the connectivity between two points for a given threshold. This information cannot be obtained by conventional variograms.

Non-centered covariance

e.g.

= 1 \*

In calculating the product of indicator variables, we must consider FOUR possibilities, of these ONLY ONE CASE will result in a value of 1 and rest all 0.

In reservoir description, knowing the connectivity or knowing the barrier (when both the values are below the threshold) is very important.

E.g. for knowing connectivity, we can define the variable as

Hence,

By substituting (a) in (b),

**Indicator variables in facies:** Indicator variables are also used for describing categorical or discrete variables, e.g. geological facies. We can define the indicator variable as:

Similarly, we can define the joint probability distribution as below:

In addition, we can define joint connectivity between the two locations with two different facies criteria,

This can be easily extended to Multipoint histogram (2F.b) for categorical variables.

Advantages:

* + On appropriately defining the threshold values and estimating indicator variogram at each threshold, we can examine how sample values are connected at each thresholds.
  + Indicators allows qualitative information like geological facies to be translated into an indicator function.
    - E.g. facies 1: (1,0,0) facies 2(0,1,0) facies 3: (0,0,1)

#### Normal Score Transform

### Biased Sampling

#### General Relative Variogram

#### Pairwise Relative Variogram

#### Non-ergodic Variogram

#### Non-ergodic correlogram

## Modelling Variograms

The primary purpose in estimating the variogram is to use this information to estimate values at unsampled locations.

### Modelling requirements

Two requirements for modelling:

* Use minimum number of parameters & models to model the variogram.
* Condition of positive definiteness, which ensures a unique solution is obtained during the estimation process. Positive definiteness can be satisfied in different dimensions. A model which satisfies positive definiteness in a higher dimension will definitely satisfy in lower dimensions.

### Models with Sills

Models which reach a constant value (sample variance) after a certain lag distance called **range**.

#### kriging modelsNugget-Effect Model

#### Spherical Model

#### Exponential Model

#### Gaussian Model

#### Combination Model

### Models without sills

Models that doesn’t reach a constant value within the region of interest.

#### Fractional Gaussian Noise fGw Model

#### Fractional Brownian Motion fgw Model

### Hole Effect Models

Models showing a periodic behaviour (representing cyclical geological processes)

#### Cosine Model

### Anisotropic Model

#### Geometric Anisotropy

#### Zonal Anisotropy

## Cross-variograms

### Estimation of Cross variograms

**Cross variogram for 2 random variables X and Y** can be written as below:

And meaning cross-variogram is symmetric.

Cross-covariance:

For a lag of :

When the 1st order stationarity holds

Although 2nd terms of the & are equal, however first terms are not equal.

Hence, or cross-covariance is not symmetric. In some mining applications they are observed to be significantly different. However, in most reservoir characterization applications we assume that to be symmetric.

From :

OR

due to stationarity.

If cross covariance is assumed to be symmetric i.e. :

**In practice, cross-variogram is calculated as:** ()

Similar to conventional variogram,

and

and

Replace in :

In most instances we assume that,

Thus,

Replacing the definition of non-ergodic cross-variogram,

For and to be equal,

The local cross-covariance must be equal to the global cross-covariance i.e.

The local mean of first data point in a pair is equal to the local mean of the second data point of the pair, i.e. ,

**Markov-Bayes approximation for cross correlation.**

X is the primary variable & Y is the covariable and we want to estimate the value of X.

Covariance of primary variable X:

Cross-covariance between variables X and Y:

We further assume a linear relationship between X and Y.

Also,

Thus,

If we assume the value of Y is only dependent on the value of X from the same location and no other X value, we can write as below:

Replacing in ,

Thus, **.** The equation allows us to calculate the cross-variance of X and Y once the covariance of X is modelled, only requiring variance of X and covariance of X,Y same location.

### Modelling of Cross-variograms

## Alternative methods of Spatial Relationships

### Modified Variograms

### Multipoint Histograms