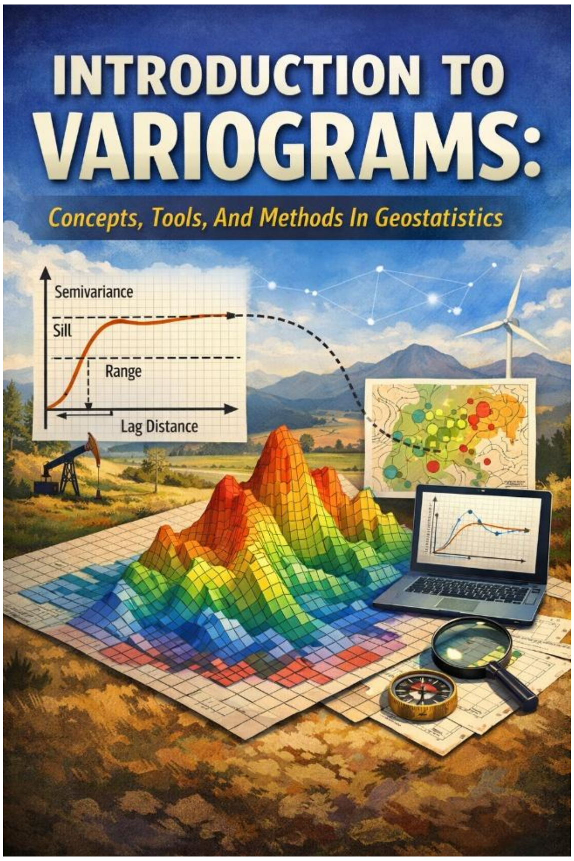
INTRODUCTION TO VARIOGRAMS: Concepts, Tools, And Methods In Geostatistics



Dr. LABBACI Yasser

Technium
2025

INTRODUCTION TO VARIOGRAMS:

Concepts, Tools, And Methods In Geostatistics

TECHNIUM

Dr. LABBACI Yasser
Researcher in Civil Engineering
University of Bechar- Tahri Mohamed
Department of Civil and Hydraulic Engineering
LMS Laboratory (Laboratoire Mécanique des Structures)
Email: Labbaci.yasser@univ-bechar.dz

Bechar, Algeria
December 2025

TECHNIUM

Dr. LABBACI Yasser
University of Bechar- Tahri Mohamed

EDITORIAL OFFICE
TECHNIUM
E.U.

For citation purposes, cite as indicated below:

LABBACI Yasser. *INTRODUCTION TO VARIOGRAMS: Concepts, Tools, And Methods In Geostatistics* ; TECHNIUM: Constanta, Romania, 2025.

FIRST EDITION 2025

doi:10.47577/book13419

Cover image courtesy of AI ChatGPT.

About the Author

Dr. LABBACI Yasser is a lecturer and researcher in Civil Engineering at the university. He holds a PhD in his field of specialization, and his research activities fall within the domain of civil engineering, with particular interest in theoretical and applied approaches related to analysis, experimentation, and the behavior of materials and structures.

He is actively involved in the academic supervision of students and contributes to the advancement of scientific research through academic publications and presentations at national and international scientific events.

Through this book, the author draws on his teaching and research experience to provide a rigorous and pedagogical synthesis, primarily intended for graduate students and researchers, with the aim of facilitating the understanding of fundamental and advanced concepts of variography in geostatistics.

Contact: Labbaci.yasser@univ-bechar.dz

Preface

Over the past decades, geostatistics has emerged as a fundamental discipline for the analysis and modeling of spatial phenomena across a wide range of scientific, industrial, and environmental applications. Understanding spatial variability, identifying trends, predicting missing values, and managing uncertainty are key challenges that researchers and practitioners face on a daily basis.

At the heart of geostatistics lies the variogram, a core tool that provides robust answers to these challenges. It plays a crucial role in quantifying spatial dependence, characterizing the continuity and heterogeneity of studied variables, and developing reliable models for estimation and simulation. Proficiency in variography enables the transformation of discrete observations into meaningful and actionable knowledge, thereby supporting informed decision-making in complex and strategic contexts.

This book is designed as a comprehensive and progressive introduction to variography. It guides the reader through essential concepts, theoretical foundations, and practical methodologies required for the effective understanding and application of variograms. By combining scientific rigor with a pedagogical approach, it presents fundamental principles, mathematical formulations, variogram properties and models, as well as methods for experimental computation and analysis.

This preface is intended for readers seeking to develop a solid understanding of variography, whether in academic research or professional practice. It aims to stimulate curiosity, foster interest in geostatistics, and illustrate how this powerful methodological framework can serve as a key tool for spatial analysis, modeling, and prediction.

Through this book, readers will gain insight not only into the theoretical foundations of variography, but also into its practical applications, highlighting geostatistics as a discipline that is both scientifically rigorous and operationally relevant, capable of addressing real-world and complex problems.

Dr. LABBACI Yasser
University of Bechar - Tahri Mohamed
Bechar, Algeria

Table of contents

1. INTRODUCTION ***08***

2. ESSENTIAL NOTIONS AND KEY CONCEPTS ***08***

2.1. Support *08*

2.2. Additivity et accumulation *09*

2.3. Standard statistics *12*

3.* CONCEPTS AND FORMULATION OF THE VARIOGRAM** ***13

3.1. Definition of the Variogram *13*

3.2. Mathematical Formulation of the Variogram *14*

4.* THEORETICAL PROPERTIES OF THE VARIOGRAM** ***15

4.1. Range and Zone of Influence *16*

4.2. Behavior at the Origin *17*

4.3. Anisotropies *19*

4.4. Presence of a Drift *21*

4.5. Nugget Effect *22*

4.6. Nested Structures *23*

5. VARIOGRAM MODELS ***24***

5.1 Permissible Models *24*

5.2. Common Models *25*

6.* CALCULATION OF THE EXPERIMENTAL VARIOGRAM** ***29

6.1. Calculation Procedure *29*

6.2. Calculation Examples *29*

6.3. Graphical Visualizations of the Variogram *29*

7.* VARIOGRAM ANALYSIS PROCEDURE.** ***34

7.1. Steps to Follow *34*

7.2. Data Verification *34*

7.3. Methodological Choices *37*

8. CONCLUSION ***38***

9. BIBLIOGRAPHIC REFERENCES ***39***

1. Introduction

In today's world, where the spatial variability of natural and industrial phenomena directly shapes scientific and technical decision-making, geostatistics has established itself as an essential analytical framework. Gaining insight into how values are distributed across space, identifying areas of continuity or disruption, and anticipating future variations are now critical challenges across a wide array of domains, including natural resource management, environmental studies, engineering, and spatial planning.

At the core of geostatistics lies the variogram, a fundamental tool that translates discrete observations into a coherent understanding of the spatial structure of phenomena. It highlights the relationships among data points, quantifies spatial dependence, and guides the development of robust models for estimation and simulation. Mastery of the variogram allows practitioners to transform raw data into actionable knowledge, integrating mathematical rigor with practical interpretation. This document presents a structured and progressive introduction to variography, blending theoretical foundations with practical applications. It addresses:

- ✓ the essential concepts underlying any geostatistical analysis,
- ✓ the definition and mathematical formulation of the variogram,
- ✓ the theoretical properties and models used to characterize spatial variability,
- ✓ and methods for computation and experimental analysis to ensure accurate and reliable data interpretation.

The aim is to provide students and practitioners with a clear, comprehensive guide for understanding, analyzing, and modeling spatial phenomena, while cultivating their ability to apply variography effectively and rigorously. By exploring these concepts, readers will discover how to convert observations into knowledge and how geostatistics can serve as a powerful tool for decision-making in complex scientific and industrial contexts.

2. Fundamental Notions and Key Concepts

2.1. Support

In geostatistics, the term support refers to « **the size** », « **shape** », or « **volume** » over which a measurement is taken or integrated. The support directly influences the statistical properties of a variable, particularly its mean and, more importantly, its variance.

Illustrative example:

Drill cores and mining blocks have vastly different volumes:

- ✓ a few kilograms for a **drill core**,
- ✓ several hundred tonnes for a **mining block**.

Although the average grade is theoretically identical at the deposit scale, the observed variance is much higher for point measurements (**drill cores**) than for larger volumes (**blocks**), because larger supports naturally smooth out spatial variability, Figure 1.

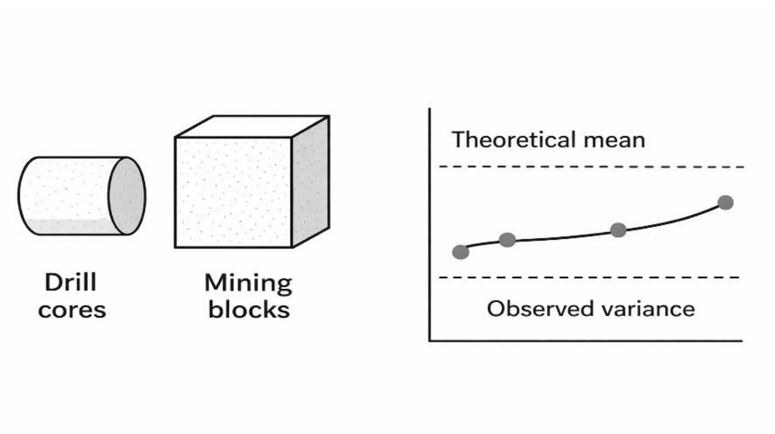


Figure 1: *Regularly Spaced Samples*

Similarly, two drillings—one conducted with a diamond drill and the other by percussion—may have identical diameters while arising from fundamentally different sampling conditions. Drilling techniques, material recovery methods, and the degree of ore disturbance vary significantly between methods. Consequently, although based on a similar geometric support, the resulting samples can exhibit markedly different statistical properties (Figure 2).

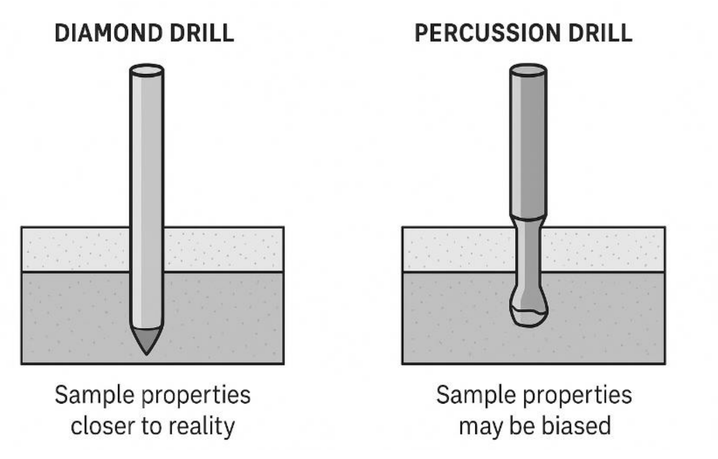


Figure 2 : *Despite similar geometric support, sampling methods can lead to markedly different statistical properties*

2.2. Additivity and Accumulation

In most geostatistical applications, the variables under analysis must be « **additive** », since additivity is essential for the validity of spatial estimates and aggregation operations. Cases where the goal is limited to a purely cartographic representation are among the few exceptions to this requirement, Figure 3.

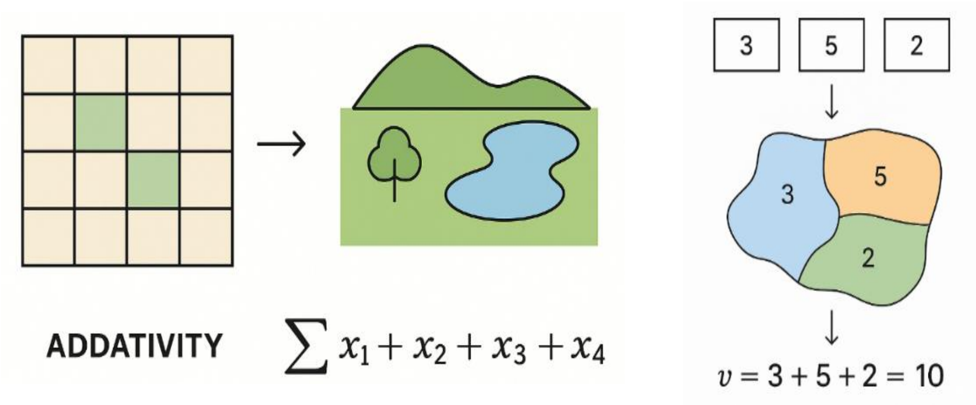


Figure 3: *Principle of additivity and accumulation*

A variable is considered « **additive** » when the value associated with a spatial entity can be obtained by summing the values measured over its constituent parts (or by computing a weighted average). In other words, for an additive variable, the average over a given area naturally corresponds to the weighted arithmetic mean of the values measured within that area, with weights proportional to the relevant volumes, lengths, or surfaces.

Exemple illustratif :

We want to estimate the average gold grade of a vein based on two drill core intersections:

- ✓ **Core 1:** length = 2 m, grade = 2 g/t
- ✓ **Core 2:** length = 3 m, grade = 10 g/t

The **average thickness** of the vein is simply the mean of the lengths:

(2 + 3) / 2 = 2,5 m

However, the **average grade** is not the arithmetic mean of the two grades:

(2 + 10) / 2 = **7,5 g/t → incorrect**

It must be calculated as a **length-weighted average**:

$$\textbf{\textit{Teneur moyenne}} = \frac{2 \times 2 + 3 \times 10}{2 + 3} = 6,8\ g/t$$

This value (6.8 g/t) is the only estimate consistent with the actual amount of metal contained in the intersected volume. The arithmetic mean (7.5 g/t) would significantly overestimate the grade of exploitable ore (Figure 4).

Accumulation represents the total amount of ore in a given area and is calculated as the product of grade and thickness. This concept is essential because it allows a realistic estimation of the exploitable grade. A simple arithmetic mean of the grades does not account for variations in thickness and can therefore lead to significant errors.

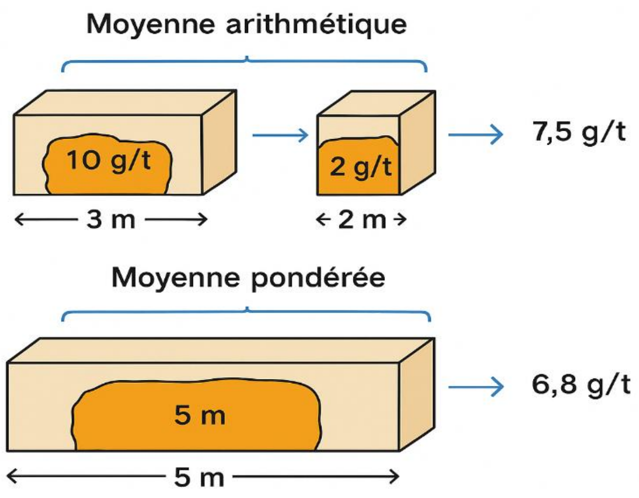


Figure 4: *Estimation of the average grade from two drill core intersections*

To avoid these errors, **accumulation** is generally used, defined as:

Accumulation = thickness × grade

The standard geostatistical approach then consists of:

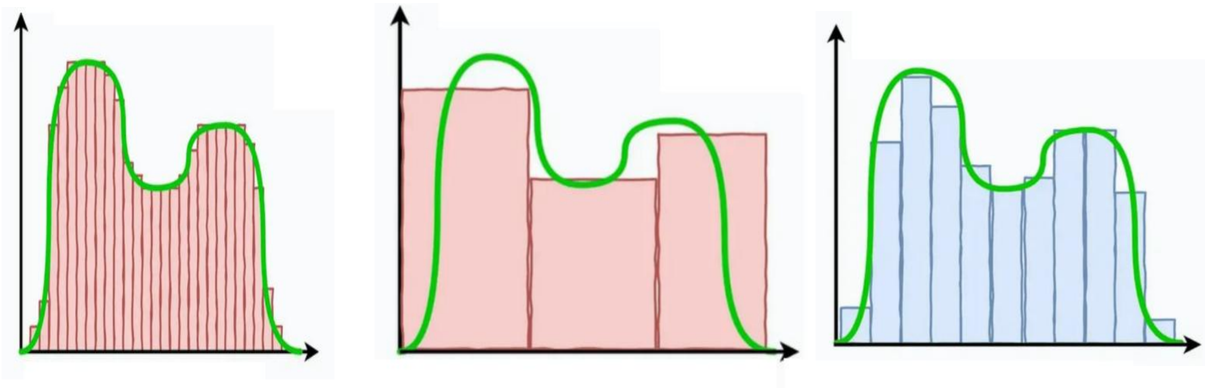
- ✓ Studying thickness and accumulation separately;
- ✓ Kriging these values to obtain accurate maps;
- ✓ Dividing the kriged accumulation by the kriged thickness to obtain the average grade.

In summary, accumulation is the key tool to link the amount of ore to its average grade, taking into account the spatial variation of thickness and density. It should be noted that weighting by thickness and density is necessary to avoid errors in estimating the exploitable grade.

2.3. Standard Statistics

Before starting the calculation of the experimental variogram, it is important to ensure that the data are consistent. A key assumption in geostatistics is that the data come from a **homogeneous population**. To do this, basic statistical measures should first be calculated: **means**, **variances**, and **correlations**. For example, the **variance** will help calibrate the sill of the variogram.

It is also essential to examine the **data histograms** to identify extreme values and check for the presence of multiple peaks. If several peaks are present, this may indicate that the data are not homogeneous, and they should then be re-evaluated (Figure 5).



(a) (b) (c)

Figure 5: *Main steps of a structural analysis in a geostatistical study*
(a): *Fine intervals;* ***(b):*** *Wide intervals;* ***(c):*** *Optimal intervals*

A **graphical representation** of the data provides a powerful tool for detecting **non-homogeneous zones** or **outlier values** that may not be apparent in histograms. By visualizing the spatial distribution or plotting the data against key variables, subtle patterns, clusters, or anomalies can be revealed.

Although performing this visual inspection can be somewhat time-consuming, it is a **critical step** in quality control, ensuring that potential errors are identified **early**. Failing to do so may result in misleading statistical analyses and could require restarting the study from scratch if significant inconsistencies are discovered at a later stage (Figure 6).

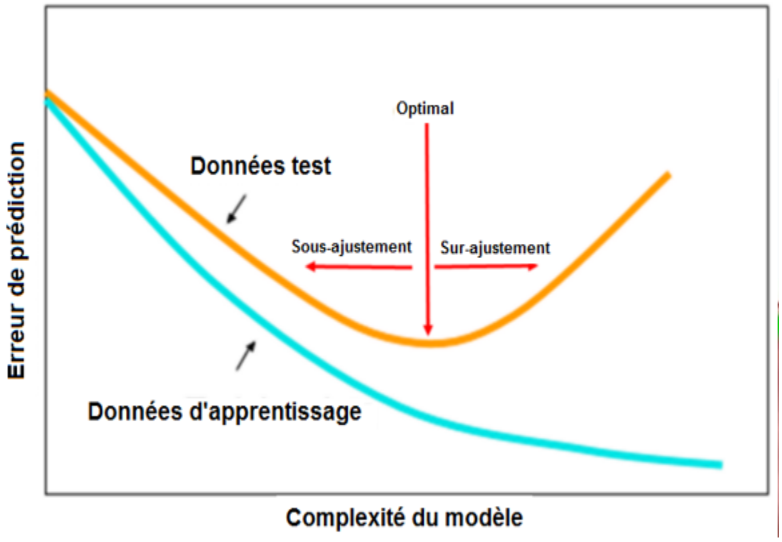


Figure 6: *Graphical representation of the data (training and test)*

3. CONCEPTS AND FORMULATION OF THE VARIOGRAM

3.1. Definition of the Variogram

The variogram is a fundamental tool in geostatistics that allows the description and quantification of the spatial dependence of a variable. In other words, it illustrates how the difference between values of a variable changes as a function of the distance separating them (Figure 7).

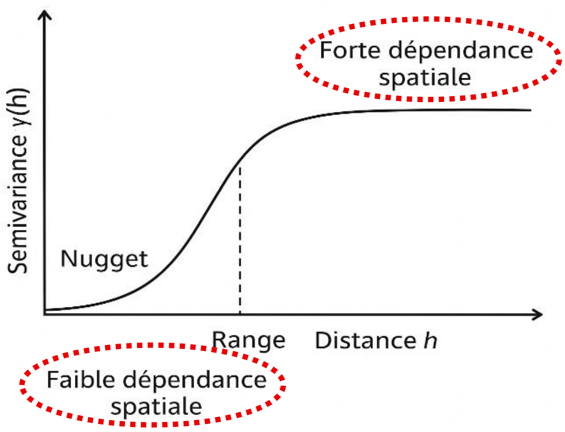


Figure 7: *Example of a typical variogram*

Thus, the variogram is an essential tool for understanding the spatial structure of the data and serves as the basis for geostatistical modeling and interpolation. In practice:

- ✓ When values at **nearby points** are very similar → **semivariance is low** → variogram is low.
- ✓ When values at **distant points** differ strongly → **semivariance is high** → variogram is high.

3.2. Mathematical Formulation of the Variogram

For a variable Z(x)Z(x)Z(x) measured at different locations x, the variogram γ(h)\gamma(of an intrinsic random function is defined by the following equation:

$$\boldsymbol{\gamma(h) = 0,5\, Var\, [Z(x+h)) - Z(x)]}$$

Where:

- ✓ $\boldsymbol{x\ et\ (x+h)}$: are points in an **n-dimensional space**;
- ✓ $\boldsymbol{n}$: can be « **1** », « **2** » ou « **3** »;

Since we have assumed that the mean of $\boldsymbol{Z(x+h)) - Z(x)}$ is zero, γ(h)\ is the mean of the squared difference, that is:

$$\boldsymbol{\gamma(h) = 0,5\, E\, [Z(x+h)) - Z(x)]^2}$$

Example:

When **n=2** (i. e., in **the plane** ′) ⇒ **x** is the point $(\mathbf{x_1}, \mathbf{x_2})$ and **h** is a vector $(\mathbf{h_1}, \mathbf{h_2})$ ⇒ in a 2-dimensional space, the variogram is a **function of the two components** « $\mathbf{h_1}$» and « $\mathbf{h_2}$» du vector « **h** ».

Note: For a fixed direction, the variogram indicates whether values differ significantly as the distance increases.The graph of γ(h)\gamma(h)γ(h) as a function of hhh has the following characteristics:

- ✓ It passes through the **origin** (for **h=0**, $\boldsymbol{Z(x+h) = Z(x)}$**)**;
- ✓ It is generally an **increasing function** of « **h** »;

- ✓ In most cases, it rises up to a certain limit called the **sill**, then flattens out. It can also increase indefinitely (Figure 8).

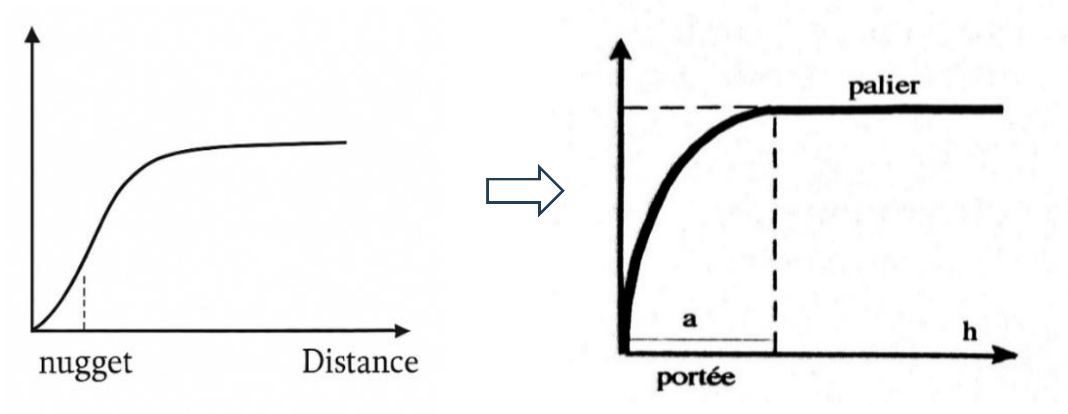


Figure 8: *Example of a typical variogram*

The variogram lies at the heart of all geostatistical methodology.
It plays a role in:

- ✓ exploratory analysis;
- ✓ model selection;
- ✓ constructing kriging equations;
- ✓ producing realistic simulations;
- ✓ detecting errors or discontinuities.

It is therefore much more than **a simple function**: it forms the foundation on which the entire spatial modeling workflow is built.

4. PROPERTIES OF THE VARIOGRAM

The variogram is a cornerstone of geostatistical analysis, serving as the primary tool for characterizing how a variable varies across space. Through both graphical representation and mathematical formulation, it captures the spatial dependence between observation points and brings to light the latent spatial structures present in the data. A solid understanding of its theoretical properties is indispensable for accurate interpretation of these structures and for the construction of sound, reliable kriging models. The sections that follow describe the principal properties of the variogram and highlight their role in geostatistical analysis and spatial modelin

4.1. Range and Zone of Influence

The more or less rapid increase of the variogram with respect to the lag distance « h » indicates the extent to which the influence of a sample decreases with distance. When the variogram reaches its upper limit (the sill), there is no longer any correlation between samples separated by that distance. This critical distance is called « **the variogram range** ». Beyond the range, the value of the variogram is exactly equal to « **the variance** » of the population. Indeed, since there is no longer any correlation between **Z(x)** and **Z(x+h)** , we then have:

$$\boldsymbol{\gamma(h) = 0,5\,Var\ \ [Z(x+h) - Z(x)]\,\gamma(h)}$$

$$\boldsymbol{\gamma(h) = 0,5\,Var\ \ \left[Var\left(Z(x+h)\right) + Var\left(Z(x)\right)\right]}$$

$$\boldsymbol{\gamma(h) = \sigma^2}$$

Not all variograms reach a sill. Some, like the one shown on the right side of Figure 9, continue to grow indefinitely with **« h »**. This highlights a fundamental difference between the variogram and the covariance: the covariance is defined only for stationary variables and is always bounded by the variance of the random function.

Note: The range is not necessarily the same in all directions, as the phenomenon may exhibit anisotropy. Moreover, even along a given direction, there can be more than one range. This occurs when multiple structures overlap, each operating at different scales. Such nested structures are referred to as **nested structures**. Examples of anisotropy and nested structures will be presented later.

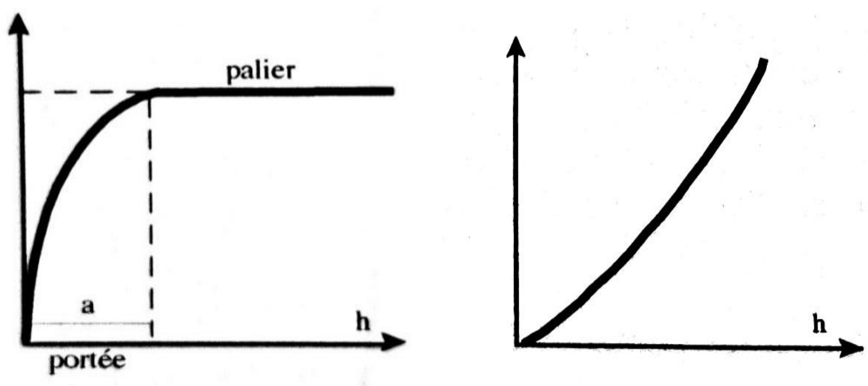


Figure 9: *Examples of Bounded and Unbounded Variograms*

4.2. Behavior Near the Origin

The behavior of the variogram near the origin provides critical insight into the variability of the measured values over very short spatial scales. Understanding this behavior is essential for distinguishing inherent natural variability from fluctuations caused by measurement errors or experimental artifacts. While previous analyses often focus on the variogram at larger distances, examining its behavior for small lag distances (« h ») reveals important information about the spatial continuity, smoothness, and micro-scale structure of the variable under study. Figure 10 and Table 1 summarize four characteristic patterns typically observed near the origin, highlighting how the variogram increases from zero and how quickly spatial correlation diminishes at very short distances.

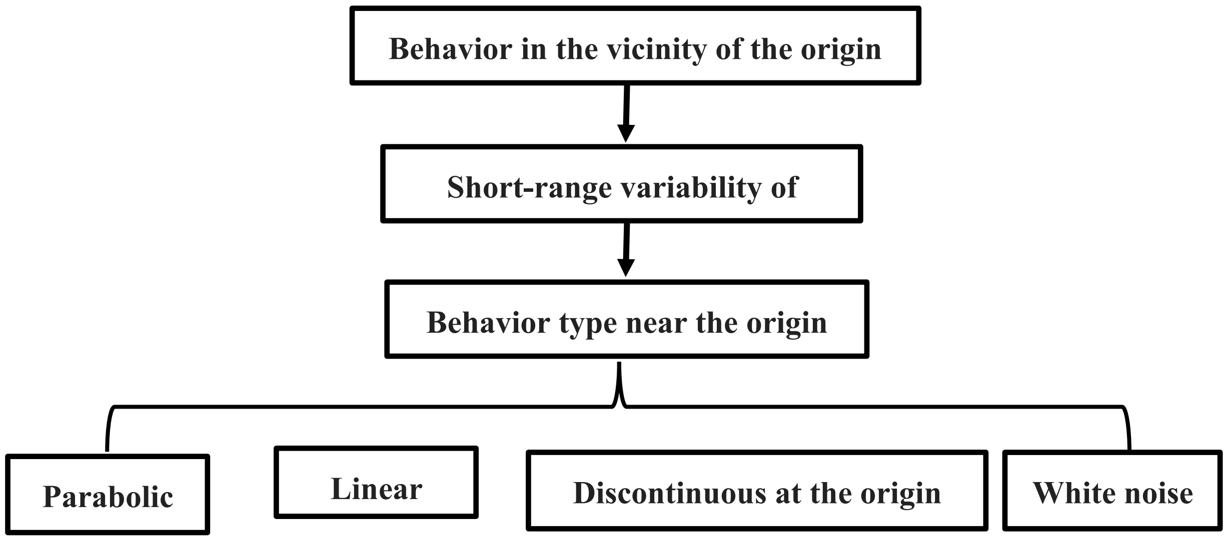
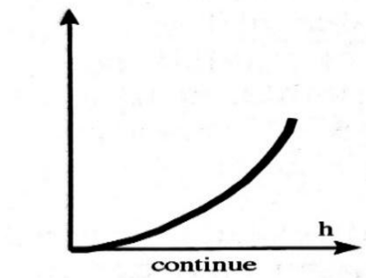
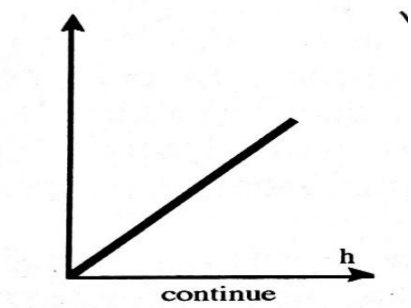
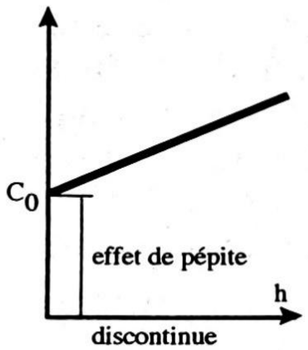


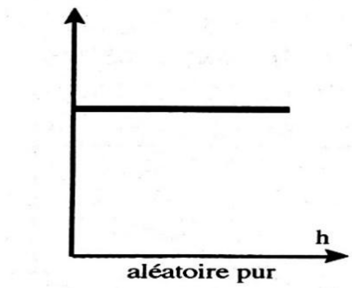
Figure 10 : *Variogram behavior at large distances*

Four characteristic patterns commonly observed near the origin:

- ✓ **Nugget effect** – A sudden jump at the origin, often reflecting measurement errors or micro-scale variability.
- ✓ **Linear behavior** – A nearly straight increase from the origin, indicating continuous but relatively rough spatial variation.
- ✓ **Parabolic or smooth rise** – Reflects a high degree of local smoothness and strong short-range correlation.
- ✓ **Plateau near the origin** – Suggests minimal variability at very short distances, implying highly correlated neighboring values.

Table 1: *Types of behavior near the origin*

Type of behavior near the origin	Definition	Schematics
Parabolic	Such behavior indicates that « **the regionalized variable** » is **highly regular**: it is **continuous** and **differentiable**. $\boldsymbol{\gamma(h) \approx h^2}$	
Linear	In this case, « **the regionalized variable**» is «**continuous**» but «**non-differentiable**», meaning it is **less regular** than in the parabolic variogram case. $\boldsymbol{\gamma(h) \approx h}$	
Nugget effect: **Discontinuous at the origin**	This type of behavior indicates that **the variable is extremely irregular** even at short distances. Values jump abruptly from zero outside the nugget to high values inside it. **$\gamma(h)$ does not tend toward «zero» as «h» approaches «zero».**	

Flat: Pure randomness or white noise	The values measured on two distinct samples show no correlation, regardless of the distance separating them. This represents the limiting case of a complete absence of structure. It is the model implicitly adopted for **multivariable regression-type analysis.**	

4.3. Anisotropies

When the variogram is calculated for all pairs of points along certain directions, such as «**North-South**» or «**East-West** », it sometimes reveals differences in behavior (i.e., **anisotropy**). If no such differences occur, the variogram then depends solely on the distance between points, which is referred to as «**isotropy**». Two main types of anisotropy can be distinguished:

4.3.1. Geometric Anisotropy: Elliptical Anisotropy

In this case, the anisotropy can be corrected through an affine transformation of the coordinates. The variation of the range and slope with **direction** can be represented, as shown in Figure 11. If the curve forms an ellipse (in 2D), this is referred to as **geometric (or elliptical) anisotropy**. Here, a simple coordinate transformation can convert the ellipse into a circle, reducing the problem to the isotropic case, as illustrated in Figure 12.

If the variogram equation in direction 1 is δ1(h)\delta_1(h)δ1(h), the overall variogram after correction takes the form:

$$\boldsymbol{\gamma(h) = \gamma_1\sqrt{(h_1)^2 + k^2(h_2)^2}}$$

Where:

- ✓ **K:** the anisotropy coefficient

Let :

$$\boldsymbol{k = \frac{port\acute{e}e_1}{port\acute{e}e_2}} \qquad\qquad \boldsymbol{k = \frac{pente_1}{pente_2}}$$

Note:

When calculating a variogram, it is important to use at least four directions. If the variogram were calculated using only two perpendicular directions, there is a risk of overlooking anisotropy, as illustrated by the right-hand example in Figure 11 and 12.

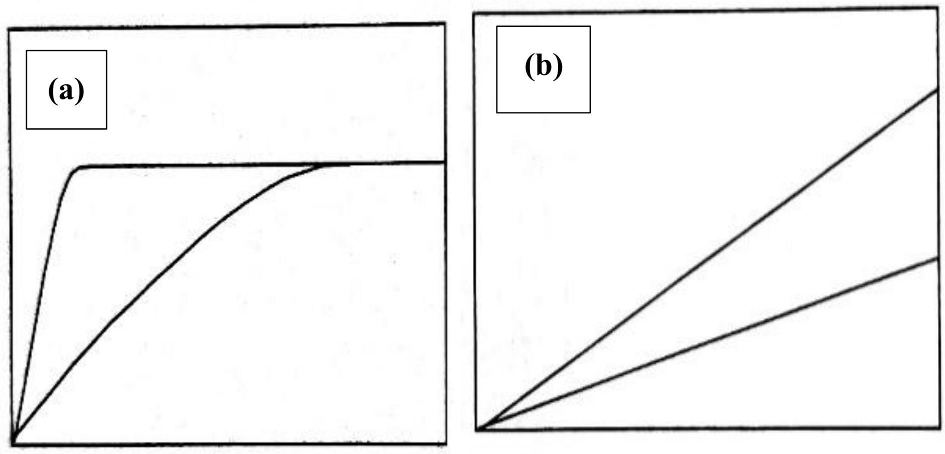


Figure 11: ***Two Typical Variogram Cases***

(a) *Effect of Changing the Range* ***(b)*** *Effect of Changing the Slope*

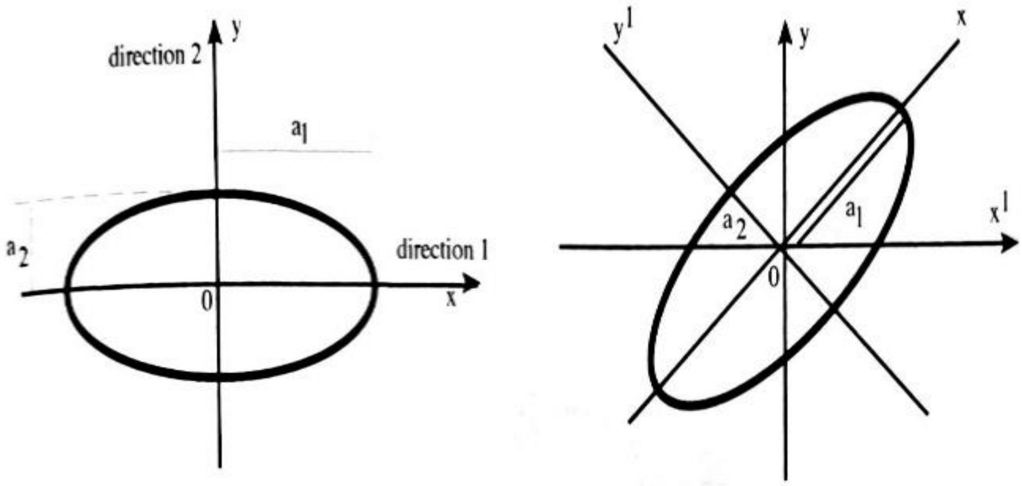


Figure 12: *Two examples of anisotropy ellipses*

4.3.2. Zonal Anisotropy: Stratified Anisotropy

There are more complex types of anisotropy. For example, in 3D, the vertical direction often plays a particular role because there is more variation between strata than within them, as shown in Figure 13.

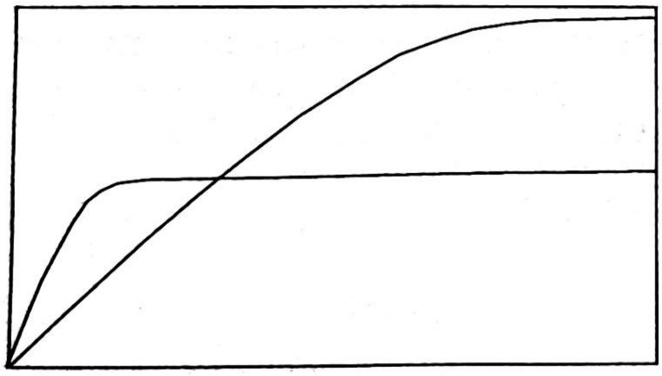


Figure 13: *Example of zonal anisotropy*

In such cases, the common practice is to separate the variogram into two parts: the first being **isotropic**, and the second depending solely on the **vertical component**:

$$\boldsymbol{\gamma(h) = variogramme\ isotrope\ + composante\ verticale}$$

$$\boldsymbol{\gamma(h) = \gamma_1\sqrt{{h_1}^2 + {h_2}^2 + {h_3}^2} + \gamma_1(h_3)}$$

4.4. Presence of a Trend

Theory shows that at very large distances, the variogram should grow more slowly than a parabola. However, in practice, one often encounters experimental variograms that increase faster than « h^2». This indicates the presence of a trend, as illustrated in Figure 14.

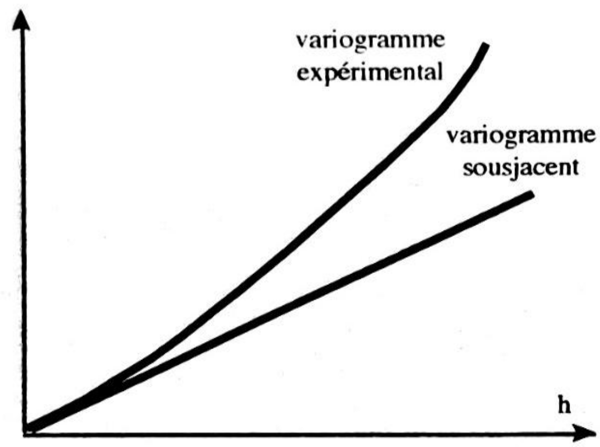


Figure 14: *Shape of a variogram in the presence of a trend*

More specifically, we have:

$$\frac{\gamma(h)}{h^2} \to 0 \qquad Quand \quad h \to 0$$

The experimental variogram (or raw variogram) is an estimate of:

$$0,5\, E\big(Z(x+h) - Z(x)\big)^2$$

It may differ from the true (or underlying) variogram, which is given by:

$$Var\,[Z(x+h) - Z(x)]$$

These two variograms coincide only if the increments have a **zero mean**. Otherwise:

$$Variogramme\ brut = variogramme\ sous - jacent\ + (biais)^2$$

$$E\big(Z(x+h) - Z(x)\big)^2 = Var\,[Z(x+h) - Z(x)] + E\,[Z(x+h) - Z(x)]^2$$

Consequently, when a trend is present, the experimental variogram yields a function that lies above the underlying variogram. To illustrate this, consider what happens with a linear trend in « **1D** »:

$$\gamma_{brut}(h) = \gamma_{réel}\,(h) + \frac{1}{2}a^2h^2$$

A quadratic term is added to the true variogram. For small values of « **h** », it is negligible, but it becomes dominant as « **h** » increases, which explains the rapid growth. Although this demonstration was carried out in « **1D** », the result also holds for linear trends in « **2D** » and « **3D** ».

4.5. Proportional Effect

A variogram is said to exhibit a **proportional effect** when its value (and particularly its sill) is proportional to the square of the local mean grade. This effect occurs with data whose histogram follows a lognormal distribution. Variograms corresponding to different areas have the same shape, but the sill of the variograms for high-grade areas is higher than that of low-grade areas, as illustrated in Figure 15.

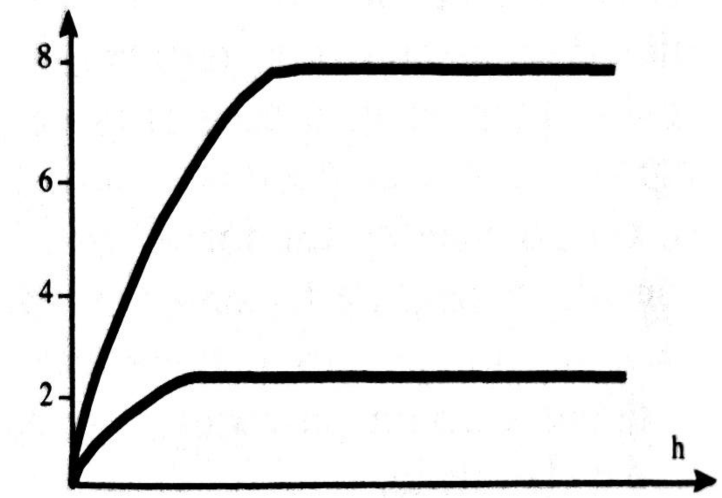


Figure 15: *Shape of a variogram in the proportional effect*

4.6. Nested Structures

A nested structure refers to the superposition of variograms with different ranges, as shown in Figure 16. This reflects a phenomenon operating at multiple scales. For example, one can identify the following four components:

- ✓ At the sample scale, a nugget effect due to measurement error;
- ✓ At the petrographic scale ($h < 1\ cm$), variations caused by the transition from one mineralogical constituent to another;
- ✓ At the metric or decametric scale, a structure corresponding to mineralized lenses;
- ✓ At large distances, a component reflecting the geometry of large mineralized bodies (clusters of lenses).

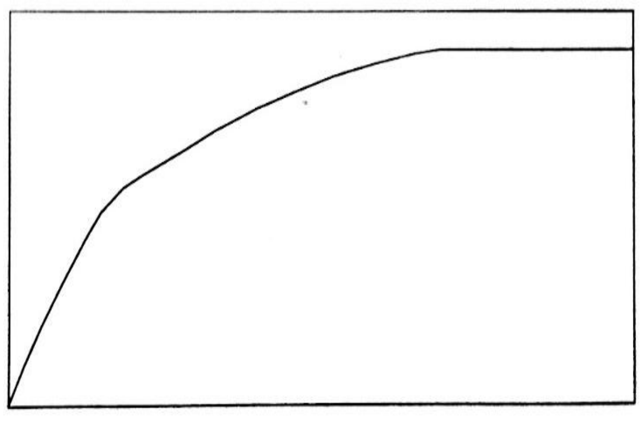


Figure 16: *Example of nested structures*

5. VARIOGRAM MODELS

5.1. Permitted Models

A list of the most common variogram models is provided below. Multiple elementary models can be combined with **positive coefficients** to obtain additional admissible models. It should be noted that determining whether a given function is positive definite is difficult, even in a one-dimensional space. The problem becomes even more complex in higher-dimensional spaces.

Example:

The piecewise linear function illustrated in Figure 17 can be used without difficulty in a one-dimensional (1D) space, as it satisfies the admissibility conditions specific to this simple case. However, when moving to two- (2D) or three-dimensional (3D) spaces, these conditions are no longer met: the function no longer guarantees the required mathematical consistency. In other words, the spatial relationships it describes can no longer correctly represent the variability of the phenomenon under study. A counterexample illustrating this lack of admissibility is presented in the work of **Armstrong and Jabin (1982)**.

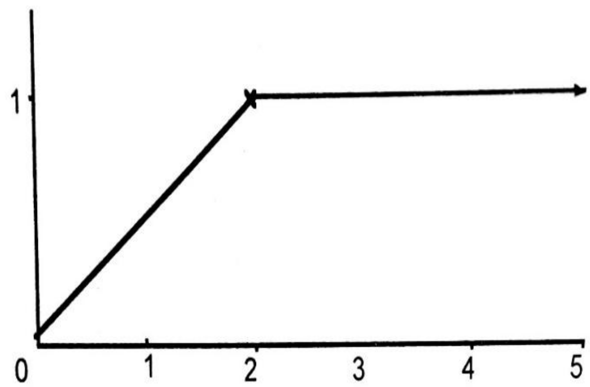


Figure 17 : *Admissibility of a piecewise linear function: Valid in « 1D », not admissible in « 2D » and « 3D ».*

to verify whether a function is positive definite—that is, whether it always ensures non-negative variances between points in space—one may follow the method proposed by Armstrong and Diamond. Although this procedure is theoretically sound, it remains difficult to apply in practice, as it involves analytically demanding calculations and checks that are hard to generalize. For this reason, in most geostatistical studies, it is recommended to rely on **covariance models** or **variogram models** whose mathematical validity has already been established. This approach helps to avoid potential problems, in particular the risk of obtaining a negative variance, which would have no physical meaning and would indicate an error in the modeling process.

5.2. Common Models

The variogram models presented below are regarded as admissible within the framework of geostatistical modeling. Models exhibiting a sill reflect the presence of a stationary structure of the random variable, characterized by a finite variance that is independent of spatial location. In contrast, unbounded variograms can only be used to describe intrinsic random variables, for which the absolute variance is not defined, but whose increments display a stable spatial organization. It should be emphasized that the list of models proposed here is not exhaustive and may be extended or adapted according to the nature of the data and the specific objectives of the study (Figure 18).

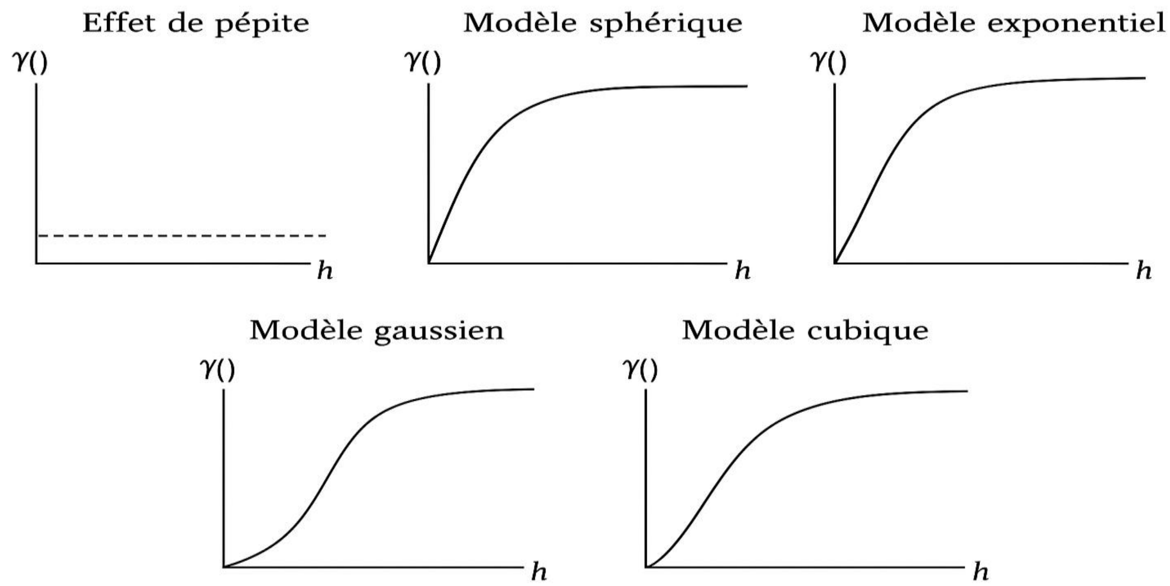


Figure 18: *Common variogram models*

Note:

Intrinsic variables are characterized by the absence of a well-defined absolute variance; however, their increments exhibit a stable spatial structure. It should be noted that the list of models presented here is not exhaustive and may be modified or extended depending on the nature of the data and the objectives of the study.

The figure above illustrates the principal variogram models commonly used in geostatistics, namely the nugget, spherical, exponential, Gaussian, and cubic models. In this representation, the horizontal axis corresponds to the separation distance h between sample locations, while the vertical axis represents the semivariance **γ(h)** (Figure 19).

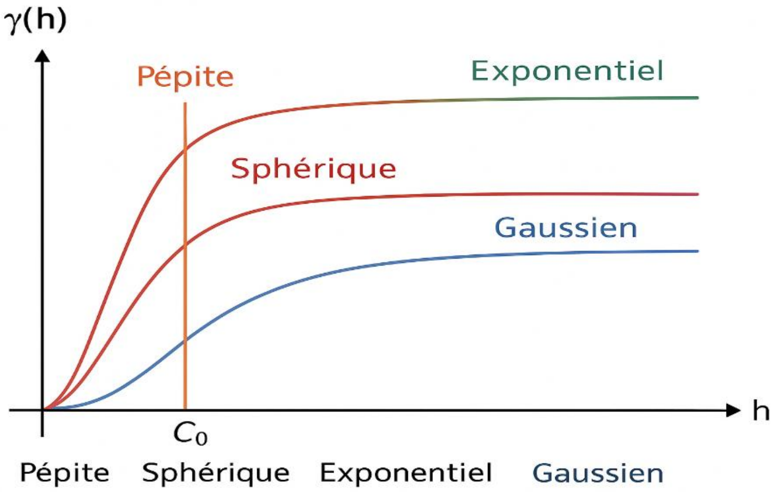


Figure 19: *Visualization of how each model captures spatial dependence and data variability*

Each model is distinguished by a characteristic shape:

- ✓ **Nugget model:** discontinuity at « h=0 », reflecting instantaneous variability or measurement errors.
- ✓ **Spherical model:** rapid increase followed by a plateau, illustrating spatial correlation limited to a specific range.
- ✓ **Exponential model:** rapid increase that approaches an asymptote, indicating a more gradual spatial continuity.
- ✓ **Gaussian model:** very smooth initial increase, suitable for continuous and smooth data.
- ✓ **Cubic model:** moderate initial rise followed by a plateau, useful for describing intermediate variations.

This figure allows visualization of how each model captures spatial dependence and data variability, thereby facilitating the selection of the most appropriate model for geostatistical analysis.

Table 2: *Common Models – Summary and Key Characteristics*

Modele	Definition	Equation	Description
Effet de pépite	It is a model that represents a completely random phenomenon. The values are not related to each other: there is **no correlation**, even when the points are very close together.	$\gamma(h) = \begin{cases} 0 & h = 0 \\ C & \lvert h \rvert > 0 \end{cases}$	In other words, each data point is independent of the others, similar to **white noise**.
Modèle sphérique	This is the most commonly used model. It shows that **the similarity between data points** increases with proximity: At first, the variability **increases almost linearly as the distance** between points grows. Then, beyond a certain distance (called **the range**), the values **stabilize**: the points are no longer correlated.	$\gamma(h) = \begin{cases} C\left[\frac{3}{2}\frac{\lvert h \rvert}{a} - \frac{1}{2}\left(\frac{\lvert h \rvert}{a}\right)^3\right] & \lvert h \rvert < a \\ C & \lvert h \rvert \geq a \end{cases}$	This model is useful because **it effectively captures the real behavior** of many natural phenomena.
Modèle exponentiel	This model shows that nearby points **are correlated,** but that this correlation **decreases rapidly** as distance increases. The variability **rises quickly at first,** then continues to grow **more slowly**, without reaching a clear limit. It is often used to represent **irregular or highly variable** phenomena, such as rainfall, **soil permeability**, or certain environmental measurements.	$\gamma(h) = C\left[1 - \exp\left(-\frac{\lvert h \rvert}{a}\right)\right]$	The similarity **decreases rapidly** with distance, without reaching a clear limit.

Modèle Gaussien	this model represents phenomena that are **highly continuous in space.** The values of nearby points are strongly correlated, and the difference between them increases slowly at first, then more rapidly with distance, until reaching a plateau.	$(\boldsymbol{h}) = \boldsymbol{C}\left[\mathbf{1} - \mathbf{exp}\left(-\frac{\lvert \boldsymbol{h} \rvert^2}{\boldsymbol{a}^2}\right)\right]$	The similarity **decreases slowly at first**, then more rapidly; it is a very **regular model**. It is used for smooth and **consistent variables**, such as **temperature**, **ore content**, or **pollution**.
Modèle cubique	This model exhibits a **smooth initial behavior**, similar to that of the Gaussian model. The difference between points **increases progressively** with distance and then reaches a **stable** level beyond a certain limit.	$(\boldsymbol{h}) = \boldsymbol{C}[\mathbf{7}\boldsymbol{r}^{\mathbf{2}} - \mathbf{8.75}\boldsymbol{r}^{\mathbf{3}} + \mathbf{3.5}\boldsymbol{r}^{\mathbf{5}} - \mathbf{0.75}\boldsymbol{r}^{\mathbf{7}}]$ **si** $\boldsymbol{r} < \mathbf{1}$ $\boldsymbol{\gamma}(\boldsymbol{h}) = \boldsymbol{C}$ **si** $\boldsymbol{r} \geq \mathbf{1}$ With $\boldsymbol{r} = \lvert \mathbf{h} \rvert / \boldsymbol{a}$. Similar to the Gaussian model, but **slightly less smooth**; it shows a **gradual variation** that eventually **stabilizes**.	it is **smooth**, but **slightly less regular** than the Gaussian model. It is used when the data change **continuously, but not perfectly.**

Note: Each variogram model (nugget, spherical, exponential, Gaussian, cubic) has a specific shape that influences spatial estimates. It is therefore essential to select the model based on the actual structure of the phenomenon under study and the sampling scheme, as two visually similar models can lead to significantly different kriging results.

6. CALCULATION OF THE EXPERIMENTAL VARIOGRAM

6.1. Calculation procedure

The experimental variogram is calculated using the following formula:

$$\gamma^*(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i + h) - Z(x_i)]^2$$

Where:

- ✓ x_i *et* $x_i + h$ **:** are sample locations separated by a distance h ;
- ✓ $N(h)$ **:** the number of such pairs of points;
- ✓ $Z(x_i)$ **:** the value measured at location x_i**;**
- ✓ $Z(x_i + h)$**:** the value at the point located at distance h.

6.2. Calculation Examples

a) Example 1: Case of regularly spaced samples

Using the data shown in Figure 20, calculate the experimental variogram by applying the corresponding formula for the first three lags. The samples are evenly spaced every 5 meters along a straight line, which makes it easier to identify the pairs of points for each considered distance.)

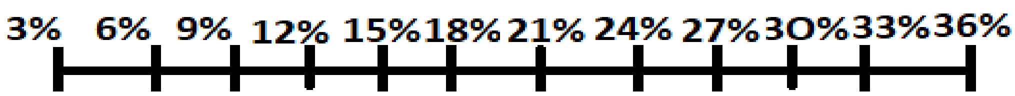


Figure 20: *Regularly spaced samples*

For the first variogram lag, we have:

$$\gamma^*(3) = \frac{1}{2 \times 12}[6^2 + 2^2 + 1^2 + 3^2 + 1^2 + 2^2 + 5^2 + 6^2 + 1^2 + 4^2 + 1^2 + 3^2]$$

$$\gamma^*(3) = \frac{1}{2 \times 12}[132]$$

$$\gamma^*(3) = 5,5(\%)^2$$

In the same way, we obtain:

$$\gamma^*(6) = \frac{1}{2 \times 11}[108] = 4,90$$

$$\gamma^*(9) = \frac{1}{2 \times 10}[99] = \mathbf{4,95}$$
$$\gamma^*(12) = \frac{1}{2 \times 11}[101] = \mathbf{5,61}$$
$$\gamma^*(15) = \frac{1}{2 \times 11}[57] = \mathbf{3,56}$$
$$\gamma^*(18) = \frac{1}{2 \times 11}[122] = \mathbf{8,71}$$
$$\gamma^*(21) = \frac{1}{2 \times 11}[38] = \mathbf{3,16}$$
$$\gamma^*(24) = \frac{1}{2 \times 11}[68] = \mathbf{6,8}$$
$$\gamma^*(27) = \frac{1}{2 \times 11}[69] = \mathbf{8,6}$$
$$\gamma^*(30) = \frac{1}{2 \times 11}[17] = \mathbf{2,83}$$
$$\gamma^*(33) = \frac{1}{2 \times 11}[34] = \mathbf{8,5}$$
$$\gamma^*(36) = \frac{1}{2 \times 11}[4] = \mathbf{0,19}$$

b) Example 2: Case of missing data

Suppose that one of the data points (**18%**) is not available. The values are now:

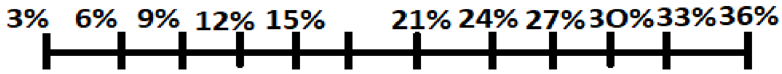


Figure 21: *Case of a missing sample*

To recalculate the experimental variogram, pairs of data containing the missing sample are ignored. Thus, for the first lag calculation, 10 data pairs are used instead of 12, as was the case previously.

c) Example 3: Case of extreme values

Suppose that an abnormally high value of 18% instead of 8% is observed in the sample. Recalculate the experimental variogram taking this high value into account, and then plot it. This will illustrate the impact of an extreme value on the construction of the variogram and demonstrate how such a value can influence the analysis results.

- ✓ **In two dimensions**

When the data are in **two dimensions** (2D), it is essential to calculate variograms along at least four different directions to assess anisotropies. If the data are arranged randomly, variograms are computed for different combinations of directions and distances. This allows obtaining the average variogram values for each group of directions and distances, thus providing a better assessment of the spatial structure.

- ✓ **In three dimensions**

The method described previously can be generalized to a **three-dimensional space** (3D) by considering classes of solid angles. However, in practice, the third dimension plays a particular role, notably due to the stratification of natural phenomena. Indeed, variations are often more pronounced in the vertical direction than in the horizontal directions. Therefore, it is more relevant to calculate variograms within the stratification plane using the method described in the previous paragraph, and then examine them in directions perpendicular to this plane. In most cases, the vertical variogram is first calculated from borehole data, followed by the computation of variograms in different horizontal directions. The first case study, presented in the following chapter, illustrates this approach.

6.3. Graphical Visualization of the Variogram

In this section, a complementary approach for representing the experimental variogram is presented. Recall that, in the classical approach, pairs of data points are grouped into distance classes (lags), and for each class the average of the squared differences between observed values is calculated. These averages define the points of the experimental variogram as it is commonly constructed.

The variogram cloud method, introduced by Chauvet (1982), provides a more detailed and informative view of the spatial structure. Rather than reducing each distance class to a single mean value, this method displays individually all squared differences associated with the pairs of points within that class. The resulting graphical representation forms a “cloud” of points, which captures not only the general trend of the variogram but also the internal variability within each lag. This representation yields a set of points scattered around the mean variogram curve, known as the variogram cloud. It allows an explicit visualization of data dispersion and of the variability inherent to each distance class. By revealing the magnitude of fluctuations, the potential presence of outliers, and the degree of data homogeneity, the variogram cloud provides a richer and more nuanced interpretation of the spatial structure. Consequently, it constitutes a valuable tool for assessing the adequacy of the chosen variogram model and for refining the overall geostatistical analysis, Figure 22.

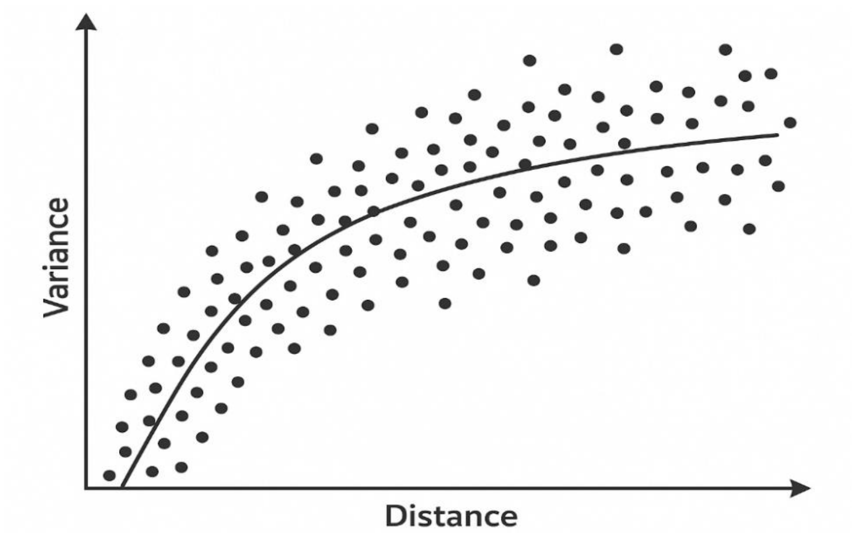


Figure 22: *Example of the Variographic Cloud Method Illustration*

The diagram illustrates an algorithm for computing variances by distance and angular classes, presented in the form of a flowchart, Figure 23. The procedure is sequential and can be described as follows:

- ✓ **Start**: The computation process begins.
- ✓ **Data input**: The values of the studied variable, denoted **Z(Xi)**, are read.
- ✓ **Definition of classes**: Distance and angular classes are defined in order to group pairs of points according to their spatial separation and directional orientation.
- ✓ **Main loop**: For each point i, from the first up to the penultimate point **(N−1)**, the following operations are performed:
- ✓ **Computation of differences:** The separation distance $\mathbf{H_{ij}=X_i-X_j}$ between points is computed, along with the squared difference of their values $\mathbf{D_2=(Z(X_i)-Z(X_j))^2}$
- ✓ **Class update**: For each class corresponding , the number of point pairs is incremented and the value D2D^2D2 is added to the cumulative sum.
- ✓ **Variance estimation by class**: For each class, the accumulated sum of squared differences is divided by the number of point pairs to obtain the mean squared difference, which constitutes the variogram estimate for that class.
- ✓ **End**: The process terminates once the variances have been computed for all distance and angular classes.

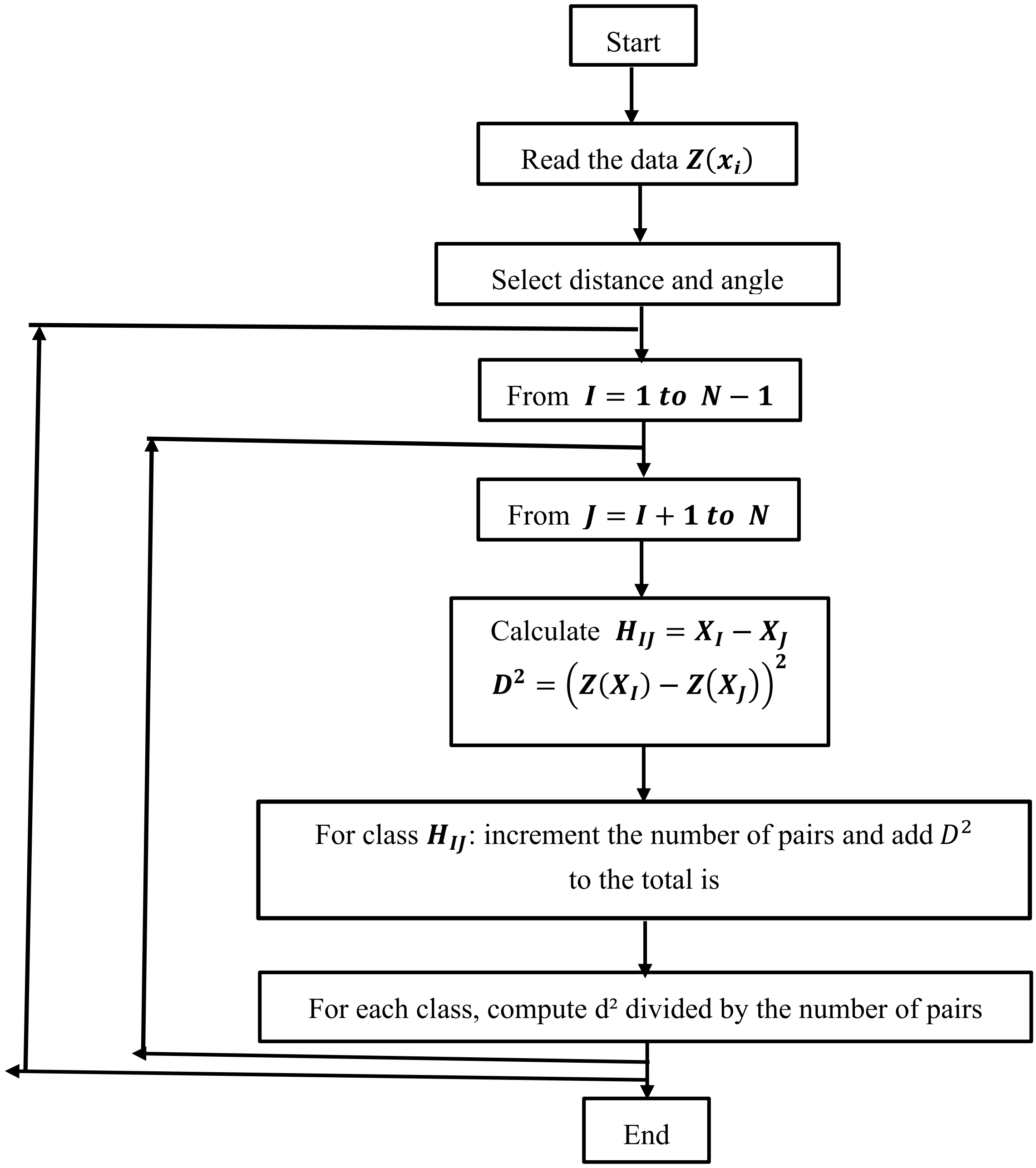


Figure 23: *Presentation of an algorithm for calculating variances by distance and angle classes (missing)*

7. VARIOGRAM ANALYSIS PROCEDURE

7.1. Steps to Follow

The first stage of any geostatistical study is structural analysis, which aims to understand how the studied variable is organized in space, Figure 24. This analysis forms the foundation of all subsequent modeling and is carried out in three main steps:

a) **Data checking and exploratory analysis:** This step involves assessing the quality and consistency of the data, identifying potential outliers, and placing the problem within its spatial and geological context. It ensures that the data are suitable for geostatistical analysis.

b) **Computation of the experimental variogram:** The experimental variogram is used to quantify spatial dependence between observations. It is the central tool of structural analysis and provides the first insights into the range and intensity of spatial variability.

c) **Fitting of a theoretical model:** A mathematical model is then fitted to the experimental variogram in order to obtain an admissible theoretical variogram. This model is essential for subsequent interpolation stages, particularly for kriging methods.

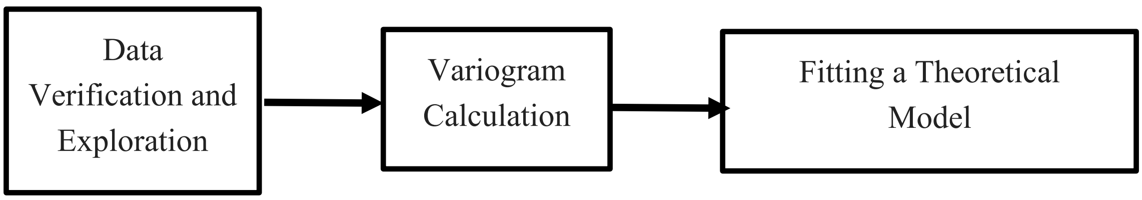


Figure 24: *Main Steps of a Structural Analysis in a Geostatistical Study*

7.2. DATA VERIFICATION AND UNDERSTANDING

The quality of a geostatistical study primarily depends on the reliability of the data and on a sound understanding of the information being used. Once the data have been collected, entered, and properly organized, it is essential to carry out a thorough verification step, which forms the foundation of all subsequent analyses.

a) Data accuracy control

The first step consists in ensuring that the data are numerically correct. This includes:

- ✓ detecting and correcting data entry or transcription errors;
- ✓ identifying missing or inconsistent values;
- ✓ verifying measurement units (mixed units, incorrect conversions, etc.);
- ✓ checking the consistency of spatial locations (duplicated, inverted, or out-of-area coordinates).
- ✓ These checks help prevent the introduction of errors that could propagate throughout the entire analysis process.

b) Assessment of representativeness

Once the data have been numerically validated, it is necessary to assess the extent to which they adequately represent the field reality. This step includes:

- ✓ computing descriptive statistics (means, variances, quantiles);
- ✓ analyzing distributions (histograms, density functions, skewness, heavy tails);
- ✓ identifying potential outliers;
- ✓ highlighting undesired structures in the data (multiple modes, trend breaks, discontinuities).

These tools make it possible to detect anomalies and to obtain an initial qualitative insight into the variability of the dataset.

c) Importance of contextual understanding

Beyond statistical checks, the success of a geostatistical study essentially depends on knowledge of the field conditions, the study objectives, and the data acquisition process. Many common errors arise from:

- ✓ a lack of understanding of the actual problem being addressed;
- ✓ insufficient recognition of the sampling process;
- ✓ omission of a major geological characteristic.

The geostatistician must therefore become familiar not only with the data themselves, but also with the real-world problem they are intended to represent.

d) Consultation with field experts

Before initiating any analysis, it is strongly recommended to meet with a geologist or an engineer who was directly involved in the exploration work. This step makes it possible to obtain crucial information, such as:

- ✓ the sampling methods used: coring, drilling, geophysical logging, surface sampling, etc.;
- ✓ any changes in protocols between different campaigns (for example, the successive use of different logging techniques);
- ✓ the geological homogeneity of the area: presence of faults, lithological variations, geological boundaries, distinct units;
- ✓ potential biases in the sampling program: oversampling of high-grade zones, lack of samples in low-grade areas, technical difficulties in certain sectors;
- ✓ operational constraints encountered in the field: limited access, variable equipment, environmental conditions.

Without this preliminary information, geostatistical interpretation is likely to be erroneous or incomplete.

e) Consequences of an inadequate initial diagnosis

Ignoring even one of these elements at the beginning of the study may lead to:

- ✓ incorrect variogram calculations;
- ✓ poorly fitted geostatistical models;
- ✓ non-representative interpolation (kriging);
- ✓ the need to redo the entire analysis at a later stage;
- ✓ a significant loss of time and reliability.

The data verification and understanding phase is therefore not a mere formality, but a **critical step that conditions the entire geostatistical analysis**.

7.3. Methodological Choices

Before starting a geostatistical study, it is essential to establish a set of methodological choices that will guide the analysis. The first step consists of precisely defining the variables of interest as well as the spatial boundaries of the study area. Once this framework is set, several fundamental decisions must be made by the geostatistical analyst, including:

- ✓ determining whether the variables exhibit **stationarity**, that is, whether their statistical properties remain stable across space;
- ✓ identifying **their support**, corresponding to the spatial unit (surface, volume, or sample) over which each measurement is collected or integrated;
- ✓ evaluating **their additive nature**, in order to decide whether it is appropriate to work directly with observed values or with their sums, averages, or integrals.

These preliminary choices form the foundation of any rigorous geostatistical study and ensure the coherence of subsequent modeling and interpolation steps.

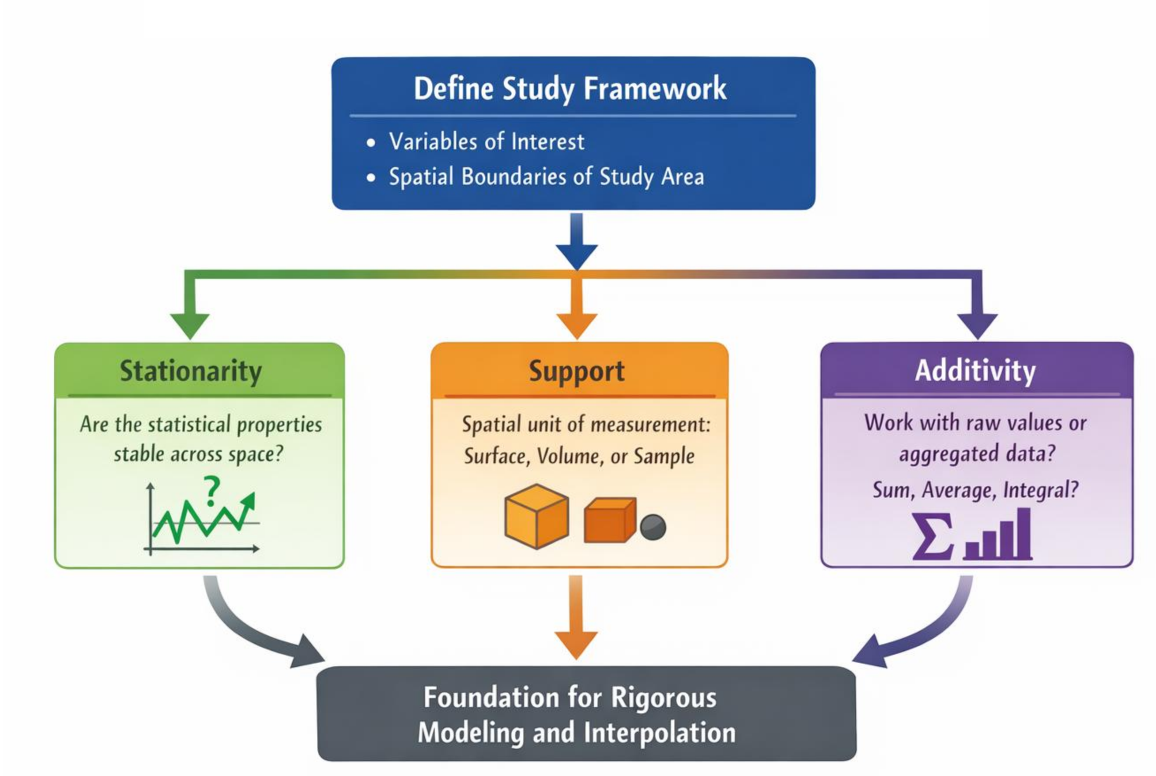


Figure 25: *Key Parameters Taken into Account by the Geostatistical Analyst*

8. CONCLUSION

At the heart of geostatistics lies the variogram, an essential tool that quantifies the spatial dependence between observations. This introductory document on variography aims to provide students and practitioners with a thorough understanding of the fundamental concepts, tools, and methods of geostatistics. Central to this approach, the variogram proves to be a key instrument for quantifying the spatial dependence of studied variables, analyzing the spatial structure of phenomena, and guiding methodological decisions in data processing.

The study of essential concepts—stationarity, support, additivity, and basic statistics—forms the necessary foundation for mastering variography. The definition and mathematical formulation of the variogram allow one to understand how variability between observations depends on distance and to translate this information into models that can be applied in kriging and geostatistical simulation. Exploring theoretical properties, such as range, nugget effect, behavior at the origin, anisotropy, and nested structures, provides the keys to fitting models and interpreting results reliably.

The calculation of the experimental variogram, accompanied by methodical analysis and graphical visualizations, enables not only the characterization of spatial variability but also the detection of anomalies, verification of data quality, and assurance of the consistency of methodological choices. Thus, variography emerges as both a descriptive and decision-making tool, essential for connecting geostatistical theory to practical field applications.

In conclusion, mastery of variography provides the ability to understand, model, and predict spatial phenomena with rigor and precision, constituting an indispensable step for any scientific, industrial, or environmental application where spatial analysis plays a strategic role. This document therefore serves as a complete and structured introduction to support learners in their journey into applied geostatistics.

9. BIBLIOGRAPHIC REFERENCES

1. Armstrong, M. (1998). *Basic Linear Geostatistics* (1st ed., XI + 155 pages). Springer.
2. Chiles, J. P., & Delfiner, P. (2012). *Geostatistics: Modeling Spatial Uncertainty* (2nd ed., 728 pages). Wiley & Sons.
3. Deutsch, C. V., & Journel, A. G. (1998). *GSLIB: Geostatistical Software Library and Users Guide* (2nd ed., 369 pages). Oxford University Press.
4. Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation* (1st ed., 483 pages). Oxford University Press.
5. Isaaks, E. H., & Srivastava, R. M. (1989). *An Introduction to Applied Geostatistics* (1st ed., 561 pages). Oxford University Press.
6. Journel, A. G., & Huijbregts, C. J. (1978). *Mining Geostatistics* (1st ed., ~600 pages). Academic Press.
7. Matheron, G. (1963). *Principles of Geostatistics* (1st ed., 115 pages). Economic Geology Press.
8. Margaret, A. (1997). *Géostatique linéaire : Application au domaine minier* (1st ed., 115 pages). Ecole des Mines de Paris.
9. Pyrcz, M. J., & Deutsch, C. V. (2014). *Geostatistical Reservoir Modeling* (2nd ed., 512 pages). Oxford University Press.
10. Wackernagel, H. (2003). *Multivariate Geostatistics: An Introduction with Applications* (3rd ed., 458 pages). Springer.
11. Emery, X., & Lantuéjoul, C. (2006). *Geostatistical Simulation: Models and Algorithms* (1st ed., 380 pages). Springer.
12. Cressie, N. (1993). *Statistics for Spatial Data* (1st ed., 900 pages). Wiley.
13. Olea, R. A. (1991). *Geostatistical Glossary and Multilingual Dictionary* (177 pages). Oxford University Press.
14. Olea, R. A. (1999). *Geostatistics for Engineers and Earth Scientists* (313 pages). Kluwer Academic.
15. Pawlowsky-Glahn, V., & Olea, R. A. (2004). *Geostatistical Analysis of Compositional Data* (181 pages). Oxford University Press.
16. Soares, A. (Ed.). (1993). *Geostatistics Tróia '92: Volume 1 & 2* (XXXII + 1089 pages). Quantitative Geology & Geostatistics.

17. Armstrong, M. (Ed.). (1989). *Geostatistics: Proceedings of the Third International Geostatistics Congress, Avignon, France* (XXIX + 1039 pages). Springer.
18. Gómez-Hernández, J., Soares, A., & Froidevaux, R. (Eds.). (1999). *geoENV II — Geostatistics for Environmental Applications* (XX + 562 pages). Springer.
19. Journel, A. G., & Rossi, R. (1989). Conditional simulation of ore bodies. *Mathematical Geology, 21*(2), 245–263. https://doi.org/10.1007/BF00896809
20. Goovaerts, P. (2000). Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology, 228*(1–2), 113–129. https://doi.org/10.1016/S0022-1694(00)00167-8
21. de Gruijter, J. J., van Groenigen, J. W., & de Kroon, H. (2014). A tutorial guide to geostatistics: Computing and modelling variograms and kriging. *Catena, 113*, 56–69. https://doi.org/10.1016/j.catena.2013.10.001
22. Zhang, H., & Deutsch, C. V. (2013). Geostatistics: A toolkit for data analysis, spatial prediction and risk management in the coal industry. *International Journal of Coal Geology, 112*, 2–13. https://doi.org/10.1016/j.coal.2012.12.002
23. Rajkai, K., & Balázs, B. (1994). Variability of variograms and spatial estimates due to soil sampling: A case study. *Geoderma, 62*(1–3), 165–182. https://doi.org/10.1016/0016-7061(94)90034-5
24. Dupont, J. (2010). *Spatial modeling of soil contamination using geostatistics* (PhD thesis). University of Paris.
25. US Geological Survey (USGS). (2001). *Geostatistical methods for mineral resource estimation* (Technical Report 01–456). USGS.
26. Remy, N., Boucher, A., & Wu, J. (2009). *Geostatistics for Environmental Scientists: SGeMS Software Manual*. Springer.
27. Goovaerts, P. (2010). Geostatistical software. In M. Fischer & A. Getis (Eds.), *Handbook of Applied Spatial Analysis* (pp. 125–134). Springer. https://doi.org/10.1007/978-3-642-03647-7_8
28. Krivoruchko, K. (2008). *Spatial Statistical Data Analysis for GIS Users*. ESRI Press.
29. Bivand, R. S., Pebesma, E., & Gómez-Rubio, V. (2013). *Applied Spatial Data Analysis with R* (2nd ed.). Springer.
30. Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data* (2nd ed.). CRC Press.
31. Gaetan, C., & Guyon, X. (2010). *Spatial Statistics and Modeling*. Springer.