



UNIVERSIDAD ESTATAL PENÍNSULA DE SANTA ELENA

FACULTAD DE CIENCIAS DE LA INGENIERÍA
ESCUELA DE PETRÓLEO
CARRERA DE INGENIERÍA EN PETRÓLEO

**“IMPLEMENTACIÓN DE ALGORITMOS GENÉTICOS PARA OPTIMIZAR
EL CÁLCULO DE VARIOGRAMAS EN TÉCNICAS GEOESTADÍSTICAS
BASADAS EN CELDAS”**

TESIS DE GRADO

Previo a la obtención del Título de:
INGENIERO EN PETRÓLEO

Presentado por:
Dennis Omar Lucín Recalde

Tutor:
Ing. Carlos Malavé Carrera

La Libertad – Ecuador
2015 – 2016

**UNIVERSIDAD ESTATAL
PENÍNSULA DE SANTA ELENA**

**FACULTAD DE CIENCIAS DE LA INGENIERÍA
ESCUELA DE PETRÓLEO
CARRERA DE INGENIERÍA EN PETRÓLEO**

**“IMPLEMENTACIÓN DE ALGORITMOS GENÉTICOS PARA OPTIMIZAR
EL CÁLCULO DE VARIOGRAMAS EN TÉCNICAS GEOESTADÍSTICAS
BASADAS EN CELDAS”**

TESIS DE GRADO

Previo a la obtención del Título de:

INGENIERO EN PETRÓLEO

Presentado por:

Dennis Omar Lucín Recalde

Tutor:

Ing. Carlos Malavé Carrera

La Libertad – Ecuador

2015 – 2016

DECLARACIÓN

Yo Dennis Omar Lucin Recalde, declaro bajo juramento que el presente trabajo es de mi autoría, además expreso que el mismo no ha sido previamente presentado para ningún grado o calificación profesional y que he dispuesto de las referencias bibliográficas incluidas en este estudio.

Por medio de la presente declaración cedo los derechos de auditoría y propiedad intelectual correspondientes a este trabajo a la Universidad Estatal Península de Santa Elena.

Santa Elena, de marzo del 2015

Dennis Omar Lucin Recalde

CERTIFICACIÓN

Ing. Carlos Malavé Carrera

TUTOR DE TESIS

Universidad Estatal Península de Santa Elena

En mi calidad de Tutor del presente trabajo, previo a la obtención del Título de Ingeniero en Petróleos elaborado por el Sr. Dennis Lucin Recalde, egresado de la Escuela de Petróleos, Facultad de Ciencias de la Ingeniería de la Universidad Estatal Península de Santa Elena. Me permito declarar que luego de haber orientado, estudiado y revisado, la apruebo en todas sus partes.

Ing. Carlos Malavé Carrera

TUTOR

AGRADECIMIENTOS

Agradezco a Dios, por darme vida y salud para llegar hasta este momento. A mis padres y hermanos por su amor y apoyo incondicional y a todas las personas que de una u otra forma han colaborado para ser de este sueño una realidad

A los directivos y personal docente de la Universidad Península de Santa Elena por la enseñanza y conocimientos impartidos y por ser entes positivos en el proceso de formación profesional. Agradezco a todos mis compañeros de estudios por compartir conocimientos, experiencias, sentimientos de alegría y compañerismo. Mi gratitud imperecedera a los Ing. Edwin Pineda, Darío Mondavi y José Ganan por sus colaboraciones y enseñanzas en el desarrollo de este proyecto.

DEDICATORIA

Dedico el presente proyecto a Dios, por darme salud para alcanzar mis objetivos propuestos, por su amor y bondad infinita hasta en los momentos difíciles, brindándome fortaleza para culminar otra etapa importante en mi vida.

A mis padres Galo y María, hermanos y familia, y amigos por su ayuda incondicional, consejos y comprensión a lo largo de toda mi vida, además por brindarme la oportunidad de estudiar y concluir satisfactoriamente mi carrera y ser los seres que han impulsado el desarrollo de este proyecto de vida.

TABLA DE CONTENIDO

DECLARACIÓN	ii
CERTIFICACIÓN.....	iii
AGRADECIMIENTOS	iv
DEDICATORIA.....	v
TABLA DE CONTENIDO	vi
LISTADO DE FIGURAS.....	x
LISTADO DE TABLAS.....	xiii
RESUMEN.....	xv
1 FUNDAMENTOS TEÓRICOS DE LA GEOESTADÍSTICA.....	1
1.1 INTRODUCCIÓN.....	2
1.2 DATOS ESPACIALES Y ANÁLISIS EXPLORATORIO.....	3
1.2.1 ESTADÍSTICA ESPACIAL	3
1.2.2 MOMENTOS DE LA DISTRIBUCIÓN.....	3
1.2.3 DESVIACIÓN ESTÁNDAR.....	4
1.2.4 VARIANZA	4
1.2.5 COVARIANZA.....	5
1.2.6 HISTOGRAMA	5
1.2.7 PERCENTIL.....	6
1.3 DATOS GEORREFERENCIADOS	7
1.4 GRÁFICOS EXPLORATORIOS	7
1.5 DEFINICIONES BÁSICAS DE GEOESTADÍSTICA.	7
1.5.1 DEFINICIÓN DE GEOESTADÍSTICA.....	7
1.5.2 VARIABLE ALEATORIA.....	9
1.5.2.1 TIPOS DE VARIABLES ALEATORIAS	9
1.5.3 VARIABLE REGIONALIZADA.....	9
1.5.4 ESTACIONARIEDAD.....	10
1.5.4.1 ESTACIONARIEDAD INTRÍNSECA	11
1.6 CORRELACIÓN ESPACIAL MUESTRAL.....	12
1.6.1 FUNCIONES DE CORRELACIÓN ESPACIAL.....	12
1.6.2 VARIOGRAMA.....	12
1.6.3 ANISOTROPÍA.....	13

1.6.3.1	ANISOTROPÍA GEOMÉTRICA	14
1.6.3.2	ANISOTROPÍA ZONAL	15
1.7	PREDICCIÓN ESPACIAL	15
1.7.1	KRIGING	17
1.7.2	KRIGINGSIMPLE	19
1.7.3	KRIGING ORDINARIO	20
1.7.4	KRIGING UNIVERSAL	20
1.7.5	KRIGING RESIDUAL	20
1.7.6	KRIGING INDICADOR	22
1.7.7	KRIGING LOG-NORMAL Y MULTI-GAUSSIANO	22
2	INTELIGENCIA ARTIFICIAL	23
2.1	INTRODUCCIÓN	24
2.2	DEFINICIÓN: INTELIGENCIA ARTIFICIAL	24
2.3	PANORAMA HISTÓRICO	26
2.4	TÉCNICAS DE LA INTELIGENCIA ARTIFICIAL	29
2.4.1	ROBÓTICA	30
2.4.2	REDES NEURONALES	30
2.4.3	PROCESAMIENTO DEL LENGUAJE NATURAL	31
2.4.4	BÚSQUEDA DE SOLUCIONES	32
2.5	COMPUTACIÓN EVOLUTIVA - ALGORITMOS GENÉTICOS.....	32
2.5.1	TEORÍA DE LA EVOLUCIÓN.....	33
2.5.2	LEYES DE LA GENÉTICA	34
2.5.3	ALGORITMOS GENÉTICOS	36
2.5.4	DEFINICIÓN DE AG	36
2.5.5	ESQUEMA DE UN AG	37
2.5.6	PARÁMETROS DE LOS ALGORITMOS GENÉTICOS.....	38
2.5.6.1	TAMAÑO DE LA POBLACIÓN	38
2.5.6.2	PROBABILIDAD DE CRUCE	39
2.5.6.3	POSIBILIDAD DE MUTACIÓN	39
2.5.7	APLICACIONES DE LOS ALGORITMOS GENÉTICOS	40
2.5.8	CODIFICACIÓN DE PROBLEMAS	41
2.5.8.1	CODIFICACIÓN BINARIA.....	41

2.5.8.2	CODIFICACIÓN NUMÉRICA.....	42
2.5.8.3	CODIFICACIÓN POR VALOR DIRECTO	42
2.5.8.4	CODIFICACIÓN EN ÁRBOL.....	42
3	LOS VARIOGRAMAS.....	44
3.1	INTRODUCCIÓN	45
3.2	DEFINICIÓN DE VARIOGRAMAS	46
3.3	ELEMENTOS DE UN VARIOGRAMA	46
3.4	CALCULO DE VARIOGRAMAS	48
3.4.1	DIRECCIONES	48
3.4.2	VALOR DE LAG.....	48
3.4.3	NÚMERO DE PARES.....	48
3.4.4	TOLERANCIA DE LAG.....	49
3.4.5	TOLERANCIA ANGULAR	49
3.4.6	ANCHO DE BANDA.....	49
3.5	CONSIDERACIONES ASOCIADAS AL CÁLCULO DE VARIOGRAMAS	50
3.5.1	INESTABILIDAD	50
3.5.2	INFLUENCIA DE LOS VALORES ATÍPICOS.....	50
3.6	MODELOS TEÓRICOS DEL VARIOGRAMA.....	51
3.6.1	EFFECTO PEPITA.....	52
3.6.2	MODELO ESFÉRICO	52
3.6.3	MODELO EXPONENCIAL	53
3.6.4	1.7.2.4 MODELO GAUSSIANO.....	54
3.6.5	MODELO CON EFFECTO AGUJERO	55
3.6.6	MODELO LINEAL	56
4	DESCRIPCIÓN DE VARIOGRAMGENETICS Y DISEÑO DEL ALGORITMO GENÉTICO.....	57
4.1	INTRODUCCION.....	58
4.2	DATOS DE INICIO DEL ALGORITMO	59
4.2.1	PARÁMETROS PARA EL CÁLCULO DE VARIOGRAMAS	65
4.2.2	PARÁMETROS DE ARRANQUE DEL ALGORITMO	66
4.3	DISEÑO DEL ALGORITMO GENÉTICO PARA EL AJUSTE VARIOGRÁFICO.....	69

4.3.1	FUNCIÓN OBJETIVO DEL ALGORITMO	72
4.3.2	OPERACIÓN DE CRUZAMIENTO.....	76
4.3.3	OPERACIÓN DE MUTACIÓN.....	78
4.4	INTERFAZ DE RESULTADOS VARIOGRÁFICOS.....	79
4.5	RESULTADOS DE LAS ELIPSES DE ANISOTROPÍAS.....	83
5	IMPLEMENTACIÓN DEL DISEÑO DEL ALGORITMO EN EL MODELAMIENTO GEOESTADISTICO.....	85
5.1	INTRODUCCIÓN.....	86
5.2	IMPLEMENTACIÓN DEL ALGORITMO PARA UN CAMPO SINTÉTICO	86
5.2.1	INFORMACIÓN GENERAL (POZOS, ARENAS).....	86
5.3	RESULTADOS DE LAS ELIPSES DE ANISOTROPÍA DEL CAMPO SINTÉTICO.	97
5.4	IMPLEMENTACIÓN DEL ALGORITMO AL CAMPO DEL ORIENTE ECUATORIANO.....	100
5.4.1	CAMPO DEL ORIENTE ECUATORIANO	100
5.4.2	LOCALIZACIÓN.....	100
5.4.3	TIPO DE AMBIENTE SEDIMENTARIO.....	101
5.4.4	INFORMACIÓN GENERAL.....	103
5.5	RESULTADOS DE LAS ELIPSES DE ANISOTROPIAS DEL CAMPO “UPSE”.....	113
6	Conclusiones y Recomendaciones	117
6.1	CONCLUSIONES.....	118
6.2	RECOMENDACIONES.....	119
	REFERENCIAS BIBLIOGRÁFICAS.....	121

LISTADO DE FIGURAS

Figura 1.1 Ejemplo de una Gráfica de Histogramas.....	6
Figura 1.2 Gráfica del Variograma Experimental.	13
Figura 1.3 Anisotropía Geométrica.	14
Figura 1.4 Anisotropía Zonal.....	15
Figura 1.5 Representación de una Predicción Espacial.	16
Figura 1.6 Representación del procedimiento kriging residual.	21
Figura 2.1 Técnicas de la Inteligencia Artificial.	29
Figura 2.2 Esquema de una Red Neuronal.....	31
Figura 2.3 Ley de la Segregación.....	35
Figura 2.4 Diagrama de un algoritmo evolutivo.....	38
Figura 2.5 Estructura de un algoritmo genético simple.	39
Figura 2.6 Individuo Genético Binario.	42
Figura 3.1 Elementos del variograma.....	47
Figura 3.2 Modelo Esférico.....	53
Figura 3.3 Modelo Exponencial.....	54
Figura 3.4 Modelo Gaussiano.....	55
Figura 3.5 Modelo con Efecto Agujero.....	56
Figura 3.6 Modelo Lineal.....	56
Figura 4.1 Interfaz Principal de VariogramGenetics.....	59
Figura 4.2 Botones de la Interfaz principal.....	61
Figura 4.3 Error de ingreso de datos 1.....	62
Figura 4.4 Error de ingreso de datos 2.....	62
Figura 4.5 Incidencia del Botón “Generar distancias y direcciones”.....	63
Figura 4.6 Interfaz Ubicación espacial de los pozos.....	64
Figura 4.7 Interfaz “Ingreso de Parámetros del Algoritmo”.....	64
Figura 4.8 Ejemplo de valores de límites para el cálculo de variogramas generado por VariogramGenetics.....	66
Figura 4.9 Ubicación del bloque Parámetros de Arranque del Algoritmo dentro de la interfaz “Ingreso de Parámetros del Algoritmo”.	67

Figura 4.10 Ejemplo de valores de parámetros de arranque del algoritmo generado por VariogramGenetics	68
Figura 4.11 Flujograma del Algoritmo Genético	70
Figura 4.12 Esquema de un individuo de la población inicial	71
Figura 4.13 Ejemplo de la acción de la función objetiva gráficamente	75
Figura 4.14 Cromosomas (individuos) codificados con parámetros a optimizar.	77
Figura 4.15 Individuos hijos resultantes del cruzamiento de un punto a nivel de bits.	77
Figura 4.16 Cromosoma con parámetros codificados.	78
Figura 4.17 Cromosoma afectado por mutación de un bit aleatorio.	78
Figura 4.18 Interfaz “Gráficas Result.”	80
Figura 4.19 Información de Avance del Algoritmo.	80
Figura 4.20 Visualización en Tiempo real del ajuste del algoritmo.	81
Figura 4.21 Partes de la sección de resultados de la interfaz “Gráficas Result.”	82
Figura 4.22 Elipses de Anisotropías creados con VariogramGenetics	83
Figura 5.1 Ubicación Espacial de los pozos del modelo sintético.	88
Figura 5.2 Límites para el cálculo de variogramas proporcionado por VariogramGenetics del modelo sintético.	89
Figura 5.3 Parámetros de Arranque proporcionado por VariogramGenetics del modelo sintético.	89
Figura 5.4 Mejor ajuste (Esférico) en el variograma 0°	93
Figura 5.5 Mejor ajuste (Exponencial) en el variograma 45°	94
Figura 5.6 Mejor ajuste (Esférico) en el variograma 90°	95
Figura 5.7 Mejor ajuste (Esférico) en el variograma 135°	96
Figura 5.8 Elipses de anisotropía del modelo Esférico y Exponencial del campo sintético.	97
Figura 5.9 Elipses de anisotropía del modelo Gaussiano y Senoidal del campo sintético.	98
Figura 5.10 Elipses de anisotropía del mejor ajuste del campo sintético.	98
Figura 5.11 Dirección preferencial de la porosidad en modelo sintético.	99
Figura 5.12 Región de Localización del campo “UPSE”	100

Figura 5.13 Sistema depositacional análogo al Miembro Arenisca U para el campo UPSE.	102
Figura 5.14 Sistema depositacional dominado por corrientes fluviales análogo al Miembro Arenisca U para el campo UPSE. Fuente:	102
Figura 5.15 Ubicación Espacial de los pozos del Campo “UPSE”.....	104
Figura 5.16 Límites para el cálculo de variogramas proporcionado por VariogramGenetics para el campo “UPSE”.....	105
Figura 5.17 Parámetros de Arranque proporcionado por VariogramGenetics para el campo “UPSE”.....	105
Figura 5.18 Mejor ajuste (Esférico) en el variograma 0° del campo “UPSE”	109
Figura 5.19 Mejor ajuste (Exponencial) en el variograma 45° del campo “UPSE”	110
Figura 5.20 Mejor ajuste (Esférico) en el variograma 90° del campo “UPSE”	111
Figura 5.21 Mejor ajuste (Exponencial) en el variograma 0° del campo “UPSE”	112
Figura 5.22 Elipses de anisotropía del modelo Esférico y Exponencial del campo “UPSE”.....	113
Figura 5.23 Elipses de anisotropía del modelo Gaussiano y Senoidal del campo “UPSE”.....	114
Figura 5.24 Elipses de anisotropía del mejor ajuste del campo “UPSE”.....	114
Figura 5.11 Dirección preferencial de la porosidad en la arenisca U del campo UPSE.	115

LISTADO DE TABLAS

Tabla 1.1 Tipos de predictores kriging y sus propiedades.	18
Tabla 2.1 Historia de la Inteligencia Artificial.....	28
Tabla 4.1 Información de Ingreso de la Interfaz Principal	60
Tabla 4.2 Ejemplo del formato de ingreso de información	61
Tabla 4.3 Ejemplo de Tabla de equivalencias de código binario con valores reales.....	71
Tabla 5.1 Información de ingreso del modelo sintético.	87
Tabla 5.2 Resultados de las optimizaciones del Algoritmo.....	91
Tabla 5.3 Información de ingreso del campo “UPSE”	103
Tabla 5.4 Resultados de las optimizaciones del Algoritmo del campo “UPSE”.	107

RESUMEN

Para poder estudiar la continuidad y regularidad espacial de las propiedades de un yacimiento, es preciso determinar la dispersión de los parámetros a estudiar y su variabilidad. Con la incorporación de toda la información disponible de la zona de estudio se realiza el proceso de caracterización de yacimientos que tiene como objetivo construir un modelo del yacimiento, lo más realista posible.

Para esto se aplica los variogramas como técnica geoestadística para la caracterización del mismo, que básicamente es una herramienta que permite analizar el comportamiento espacial de las variables regionalizadas o propiedades del yacimiento. En el cálculo de variogramas se debe definir distintos parámetros tales como direcciones y números de direcciones, azimut, tolerancia angular, ancho de banda, distancia, etc. para su posterior modelamiento.

El presente trabajo desarrollará un código fuente que permita automatizar el cálculo de estos variogramas para que de esta forma se pueda optimizar la selección de los parámetros necesarios para su ajuste en el modelamiento geoestadístico. Esta técnica consiste en la implementación de Algoritmos Genéticos (AG), probados por medio de la experimentación, que puedan ser utilizados para la optimización del cálculo de variograma basadas en celdas o malla.

La técnica de Algoritmos Genéticos (AG) explota eficientemente la información histórica (memoria implícita) del yacimiento para considerar nuevos puntos de investigación o zonas de interés para lo cual estará diseñado con la intención de mejorar el cálculo y el ajuste variográfico.

Capítulo I

1 FUNDAMENTOS TEÓRICOS DE LA GEOESTADÍSTICA

1.1 INTRODUCCIÓN

Para el estudio y análisis de datos espaciales existen una gran variedad de métodos y entre uno de ellos destaca el método geoestadístico, ya que presenta una mejor calidad de sus resultados por su carácter operativo. En este tipo de método resalta aspectos como herramientas estadísticas, métodos de simulación, krigeaje, o de simulación condicional, y otros tipos de aspectos que no dejan de centrarse en fundamentos probabilísticos comunes.

El término “geoestadística” en sí fue establecido por G. Matheron (1962), pero ésta tuvo sus inicios a principios del siglo XIX con el estudio de la variación de la producción en distintas parcelas realizado por Mercer y Hall en 1911. Ese año fue el comienzo para el desarrollo de la geoestadística ya que era la primera síntesis de la dependencia espacial, nueve años más tarde Ronald Aymer Fisher realizó un trabajo donde identificó la variación espacial y redujo los efectos de la variación en rangos cortos. Esta identificación ayudó a A. N. Kolmogorov al reconocimiento de la correlación espacial en 1930, para que de esta manera planteara una “función estructural” para describirla.

Posteriormente, comenzaron en 1937 Youden y Melich a utilizar como herramienta al análisis de la varianza, para estimar la variación espacial que años atrás Fisher había identificado, esta herramienta sirvió para el estudio de esa variación pero estimándola a diferentes distancias. Ya en los años 1940 y 1950 comenzaron a desarrollar un estudio experimental en unas minas de oro para mejorar estimaciones, lo que conlleva al desarrollo de un método basado en la estimación ponderada para resolver problemas, este método fue empleado por D. G. Krige. Finalmente George Matheron, derivó soluciones a problemas de estimación a partir de la teoría de variables regionalizadas que el fundamentó, dando por primera vez la terminología de Geoestadística como una ciencia para el análisis de datos espaciales. De

allí en adelante se han generado numerosos estudios y desarrollos como nuevas herramientas metodológicas y creación de sofisticados software que ha ayudado y aportado a la Geoestadística.

Sabiendo el surgimiento de esta disciplina es importante el conocimiento de cada uno de los términos conceptuales que abarca la geoestadística hoy en día. De esta forma este capítulo tiene un enfoque teórico-práctico, que servirá para el entendimiento del mismo, realizando una descripción que se requiera como conocimientos básicos y fundamentales de la geoestadística que se requerirán en este proyecto.

1.2 DATOS ESPACIALES Y ANÁLISIS EXPLORATORIO.

1.2.1 ESTADÍSTICA ESPACIAL

La Estadística espacial podemos definirla como una reunión de un conjunto de metodologías adaptadas para el estudio de datos que se relacionan al cálculo de variables aleatorias en varios sitios (puntos del espacio) de una zona. Es decir que la estadística espacial se enlaza con el análisis de realizaciones de un proceso estocástico estudiando fenómenos aleatorios. La estadística espacial conoce un progreso significativo en cuantiosas ramas tales como ciencias del medio ambiente y de la tierra, ingeniería minera, geografía, biología, ecología, economía, agronomía, procesamiento de imagen, etc.

1.2.2 MOMENTOS DE LA DISTRIBUCIÓN

Momentos son los promedios de las series de potencias de las desviaciones de la variable con relación a la media. El momento de primer orden es la

esperanza matemática y los momentos de segundo orden considerados en geoestadística son la varianza y la covarianza de dos variables aleatorias. (Díaz Viera, 2002)

1.2.3 DESVIACIÓN ESTÁNDAR

Es una medida de dispersión, es decir el grado en que los datos numéricos yacen a dispersarse de un valor promedio. Matemáticamente está definido como la raíz cuadrada de la suma de los cuadrados de las desviaciones en relación con la media o desviación de la media cuadrática, dicha de otra forma la raíz cuadrada de la varianza. Está representada por la siguiente ecuación:

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} \quad (1)$$

Donde:

S: Desviación estándar.

n: Numero de mediciones

$X_1, X_2 \dots X_n$: Mediciones.

1.2.4 VARIANZA

La expresión matemática nos indica que es el cuadrado de la desviación estándar (2), es decir es un valor numérico que nos muestra una representación de la dispersión de una variable.

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \quad (2)$$

Donde:

S^2 : Varianza.

n: Numero de mediciones

$X_1, X_2 \dots X_n$: Mediciones.

También se la puede expresar de la siguiente manera:

$$Var(X) = E((X - E(X))^2) \quad (3)$$

1.2.5 COVARIANZA

La covarianza mide la fuerza de la relación lineal entre dos variables, es una medida que nos muestra la variabilidad conjunta de dos variables numéricas (cuantitativas). Empleada cuando no se puede controlar una o más variables extrañas.

$$S_{xy} = \frac{\sum_{i=1}^n (X_i - \bar{X})(y_i - \bar{y})}{n} \quad (4)$$

1.2.6 HISTOGRAMA

El histograma es un tipo específico de gráfico de barras que indica la frecuencia de distribución de una variable establecida, como lo muestra la figura 1.1. Esto nos permite examinar y comparar esquemas de comportamiento en la variable, que no son aparentes a simple vista al calcular un porcentaje o la media.

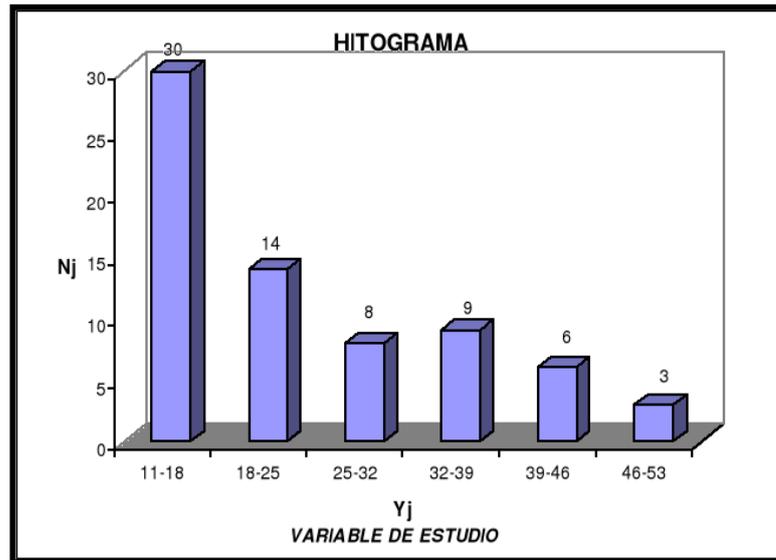


Figura 1.1 Ejemplo de una Gráfica de Histogramas.

Fuente: Mendoza Freddy

1.2.7 PERCENTIL

Expresa a la distribución de los datos en 100 partes iguales, dando 99 valores llamados percentil de una tabla de datos. De ésta se deriva los Cuartiles (Q) que son valores que dividen a la distribución de los datos en 4 partes iguales, es decir es el 25 % del mismo. Mientras los valores que dividen a la distribución de los datos en 5 partes iguales se los denominan Quintiles (q), que representa el 20 % de los datos.

$$P_k = L_i + \frac{\frac{K*N}{100} - F_{i-1}}{f_i} * a_i \quad (5)$$

Donde:

L_i : Límite inferior del intervalo donde se encuentra el percentil.

N : La suma de las frecuencias absolutas.

F_{i-1} : Frecuencia acumulada anterior a la clase del percentil.

a_i : Amplitud de la clase.

1.3 DATOS GEORREFERENCIADOS

Las aproximaciones de las características de interés en un estudio regionalizado, tienen expresamente relacionadas las coordenadas de las zonas en donde éstas fueron tomadas. Cuando la zona de estudio es demasiado grande se usa un geoposicionador para constituir dichas coordenadas. Esto nos permite la identificación de valores de la variable en su ubicación geográfica, la evaluación de la forma de la distribución, variabilidad y correlación, para que en la utilización de la teoría geoestadística nos indique los valores válidos o para definir que procedimiento de predicción es el más conveniente.

1.4 GRÁFICOS EXPLORATORIOS

Para los datos georreferenciada se pueden emplear histogramas, diagramas de tallos y hojas, con el fin de identificar ubicación, variabilidad, forma, etc. Adicionalmente los gráficos de dispersión son muy útiles tanto para la detección de relaciones entre las variables como para la identificación de tendencias en el valor promedio de la variable en la región (relación entre la variable medida y las coordenadas geográficas). (Ramón Giraldo Henao)

1.5 DEFINICIONES BÁSICAS DE GEOESTADÍSTICA.

1.5.1 DEFINICIÓN DE GEOESTADÍSTICA

La Geoestadística es una rama de la estadística aplicada que se concentra en el análisis, modelación y predicción de la variabilidad espacial de fenómenos en Ciencias de la Tierra. Matheron definió a la Geoestadística

como "la aplicación del formalismo de las funciones aleatorias al reconocimiento y estimación de fenómenos naturales" (Matheron, 1962).

Esta se caracteriza por tomar en cuenta la relación espacial de las variables en estudio. Se enfatiza a analizar, procesar y deducir resultados de datos georeferenciados. La geoestadística es una serie de técnicas que permite el análisis y predicción de valores distribuidos en el espacio y/o en el tiempo, dichos valores se asumen correlacionados entre sí.

Cuando el fin es hacer predicción, la geoestadística aplica básicamente dos fases. La primera es el análisis estructural, en ella se detalla la correlación entre puntos en el espacio y la segunda etapa se realiza predicciones en lugares de la región no muestreados a través de la técnica kriging. Este es un proceso que calcula una media ponderado de las observaciones muestrales. Los pesos asignados a los valores muestrales son apropiadamente determinados por la estructura espacial de correlación establecida en la primera etapa y por la configuración de muestreo (Petitgas, 1996). Posteriormente, el resultado de la Geoestadística muestra dos productos finales:

- Mapa (Geoestadística 2D) o Volumen (Geoestadística 3D) de valores esperados.
- Distribución local de frecuencias en cada punto del plano o del espacio.

El primero es utilizado para el modelado espacial del reservorio mientras que el segundo permite analizar el comportamiento probabilístico y de riesgo del mismo, por medio de la cuantificación de la incertidumbre asociada en cada punto del entorno analizado.

1.5.2 VARIABLE ALEATORIA

Una variable aleatoria es una función que asigna un número real a cada uno de los resultados de un experimento. Dado un experimento aleatorio y asociado al mismo, en un espacio probabilístico (E, \mathcal{A}, P) , una variable aleatoria es una aplicación $X: E \rightarrow R$, a cada valor de X del espacio muestral le corresponde un número real. Se dice que X es una variable aleatoria si para cualquier X perteneciente a R , el conjunto de los sucesos elementales le hace corresponder un valor que verifica que:

$$\forall X \in R \quad X(S) \leq X$$

1.5.2.1 TIPOS DE VARIABLES ALEATORIAS

Estas se clasifican en dos tipos discretas y continuas:

- DISCRETA: La variable aleatoria X se dice que es discreta si los números asignados a los sucesos elementales de E son puntos aislados. Sus posibles valores constituyen un conjunto finito o infinito numerable.
- CONTINUA: La variable aleatoria X será continua si los valores asignados pueden ser cualesquiera, dentro de ciertos intervalos, es decir, puede tomar cualquier valor de R .

1.5.3 VARIABLE REGIONALIZADA.

Una noción clave en Geoestadística es la variable regionalizada, que tiene propiedades intermedias entre una variable totalmente aleatoria y una plenamente determinística. Se puede definir a una variable regionalizada como una variable aleatoria cuya realización depende de la ubicación. Ejemplos típicos de variables regionalizadas son el perfil de pozo registrado

con algún objeto de sondeo, la prominencia topográfica de algún terreno, etc. En contraste con una variable totalmente aleatoria, una variable regionalizada posee cierta correlación punto a punto, pero estos cambios son tan complejos que no consiguen ser definidos por una función determinística.

Esta variable es espacialmente continua, pero por otra parte en la práctica se dispone de muestras específicas tomadas en lugares puntuales, por ejemplo en los pozos (las propiedades). La forma, orientación, tamaño y distribución espacial de estas muestras se la denominan como soporte de la variable regionalizada, si por alguna razón cambia, entonces la variable regionalizada también lo hace.

1.5.4 ESTACIONARIEDAD

Una variable aleatoria $Z(u)$, $u \in A$ se dice que es estacionaria en la región A si la función de distribución acumulada es invariante bajo cualquier traslación C efectuada sobre sus localizaciones: $F(u_1, \dots, u_k; Z_1, \dots, Z_k) = F(u_1 + C, \dots, u_k + C; Z_1 + C, \dots, Z_k + C)$, para cualquier vector de traslación C . Es decir, esta hipótesis establece el grado de homogeneidad espacial del fenómeno.

Como sólo se dispone de una realización discreta de la variable $Z(u)$, es inevitable que haya que asumir ciertas hipótesis para poder llevar a cabo algún tipo de estudio estadístico. El tipo de estacionariedad asumido indica qué tipo de inferencia estadística puede realizarse con el modelo probabilístico.

Se considera un proceso estocástico $\{Z(u), u \in D\}$, donde D es un subconjunto de R^d (espacio Euclídeo d -dimensional). El proceso Z se dice que es Gaussiano si, para cualquier natural $k \geq 1$ y las localizaciones

u_1, \dots, u_k , el vector $(Z(u_1), Z(u_2), \dots, Z(u_k))$ tiene una distribución normal multivariante.

1.5.4.1 ESTACIONARIEDAD INTRÍNSECA

Existen funciones aleatorias $Z(u)$ que representan a fenómenos físicos que muestran una capacidad casi ilimitada de variación, por lo que para estas funciones no están definidas la varianza ni la covarianza. Sin embargo existen casos en que sus crecimientos o diferencias $Z(u+h) - Z(u)$ tienen una varianza finita. En otras palabras, las diferencias son estacionarias de segundo orden.

Por lo tanto las funciones aleatorias intrínsecas son aquellas que cumplen las siguientes condiciones:

- El valor esperado de las diferencias es:

$$E[Z(u+h) - Z(u)] = 0 \quad (6)$$

- La varianza de las diferencias es:

$$Var[Z(u+h) - Z(u)] = 2\gamma(h) \quad (7)$$

Estas condiciones se conocen como Hipótesis Intrínseca, se puede ver que una función aleatoria estacionaria de segundo orden es siempre intrínseca y lo contrario no se cumple. A las funciones que cumplen con la hipótesis intrínseca se les considera como débilmente estacionarias.

1.6 CORRELACIÓN ESPACIAL MUESTRAL

1.6.1 FUNCIONES DE CORRELACIÓN ESPACIAL

La primera fase en el proceso de un análisis geoestadístico es la determinación de la dependencia espacial entre los datos medidos de una variable. Esta etapa es además distinguida como análisis estructural. Para llevarla a cabo, con base en la información muestral, se pueden usar tres funciones: El variograma, el covariograma y el correlograma.

1.6.2 VARIOGRAMA

Cuando se definió la estacionariedad intrínseca se estableció las condiciones cuando una función aleatoria es explícitamente intrínseca. A esta función denotada por $2\gamma(h)$ se la denomina variograma. El variograma es una herramienta que permite analizar el comportamiento espacial de una variable sobre un área definida, obteniendo como resultado la influencia de los datos a diferentes distancias.

La función semivarianza, es la mitad del variograma $\gamma(h)$ y caracteriza las propiedades de dependencia espacial del proceso. Dada una realización del fenómeno, la función de semivarianza es estimada por el método de momentos a través del semivariograma experimental (muchos autores denominan también a ésta como variograma experimental, figura 1.2), que se calcula mediante la ecuación:

$$\bar{\gamma}(h) = \frac{\sum (Z(x+h) - Z(x))^2}{2h} \quad (8)$$

Donde:

$Z(x)$: valor de la variable en un sitio x .

$Z(x+h)$: valor muestral separado del anterior por una distancia h

h : número de parejas que se encuentran separadas por dicha distancia.

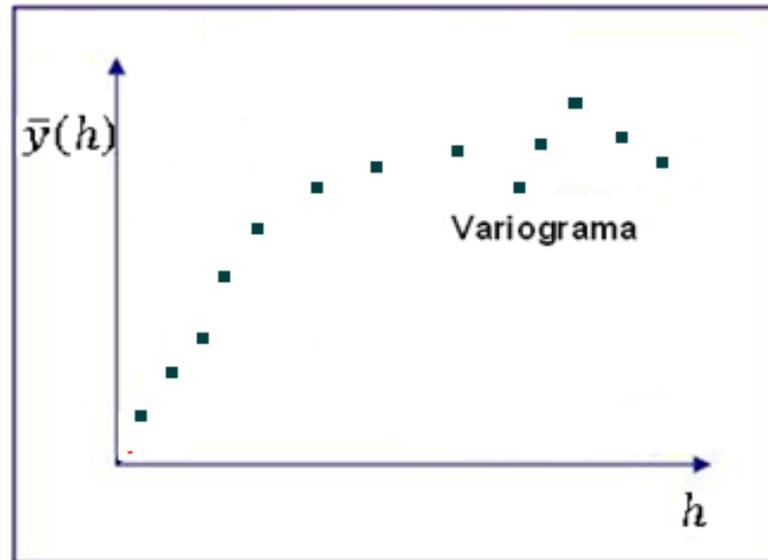


Figura 1.2 Gráfica del Variograma Experimental.

Elaborado por: Dennis Lucin

En el capítulo 3 se entrará a detalle sobre los variogramas, análisis, cálculo, elementos, parámetros de consideración para el cálculo, descripción del mismo, etc. y su posterior modelamiento, para una mejor comprensión de lo que se deberá optimizar con la aplicación de los algoritmos genéticos, que es el objetivo de ésta tesis.

1.6.3 ANISOTROPÍA

Por lo general las variables regionalizadas poseen un comportamiento anisótropo cuando demuestran direcciones específicos de variabilidad. De hecho muchos fenómenos geológicos muestran una clara anisotropía espacial referente a la continuidad. En la realidad, una anisotropía real generalmente se relaciona con la existencia de direcciones genéticas o estructurales del fenómeno, que usualmente es conocido previamente.

Para poder analizar y estudiar la presencia de anisotropías es primordial e importante que se realice el cálculo de los variogramas en diferentes direcciones, que por lo general se las calcula a 0°, 45°, 90° y 135°. De esta manera la representación conjunta de los variogramas calculados nos permitirá observar si tienen un comportamiento similar o no.

Se reconocen al menos dos tipos de anisotropías: anisotropía elíptica o geométrica y la anisotropía zonal.

1.6.3.1 ANISOTROPÍA GEOMÉTRICA

Este tipo de anisotropía se presenta cuando el alcance cambia con la dirección, mientras que la meseta o sill (elementos del variograma) se mantiene constante. De igual forma Kelkar y Pérez expresan que las anisotropías geométricas se manifiestan cuando los variogramas en las direcciones de la continuidad máxima y mínima muestran una forma y meseta similar pero diferentes rangos.

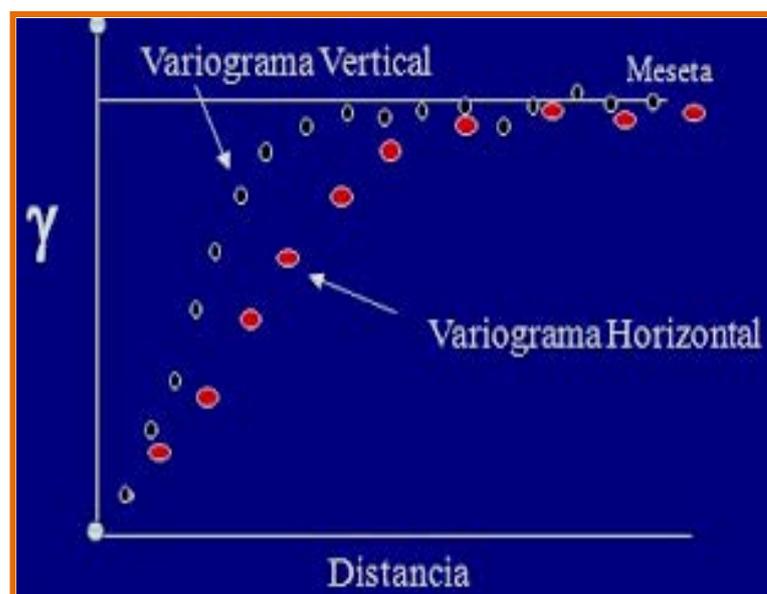


Figura 1.3 Anisotropía Geométrica.

Fuente: Petrobras

1.6.3.2 ANISOTROPÍA ZONAL

La anisotropía zonal se manifiesta en el momento que existen características iguales en el alcance pero en el sill o meseta varia en los distintos variogramas calculados a varias direcciones, tal y como se muestra en la figura 1.4. En la práctica este suele ser el más frecuente.

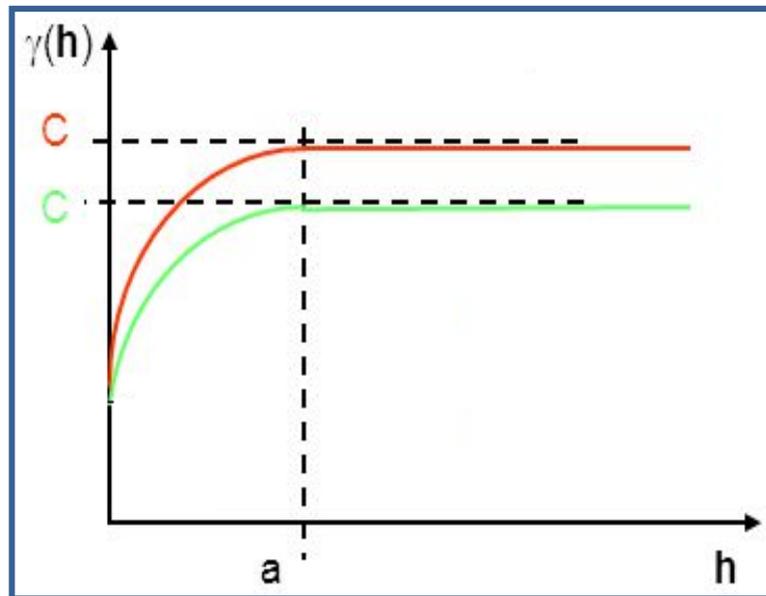


Figura 1.4 Anisotropía Zonal.

Modificado de: Eduardo G. Camargo

1.7 PREDICCIÓN ESPACIAL

Dentro de un yacimiento es seguro que no es posible conocer los valores de la variable para cada punto, dentro del área en la que se distribuye. Para eso la predicción espacial o la estimación nos permite predestinar o predecir los valores de este tipo de variables en lugares en la cual se desconoce su valor, a través de la utilización de los valores de la variable en lugares donde son conocidos (Ver figura 1.5).

La aproximación geoestadística de valoración utiliza las herramientas probabilísticas de manera que aprovechan la incertidumbre de lo sucedido en la generación del atributo o propiedad que se quiere estudiar y analizar. En un modelo probabilístico, la información disponible es vista como el resultado de un proceso aleatorio. Por ejemplo, si tratáramos de puntualizar todos los pasos involucrados en la formación de un yacimiento de petróleos, tendríamos un desconocimiento tan amplio de toda la situación, que su complejidad surgiría como la de un comportamiento aleatorio; pero esto no representa que los procesos sean en realidad aleatorios, sencillamente lo que ocurre es que se ignora o se desconoce lo sucedido.

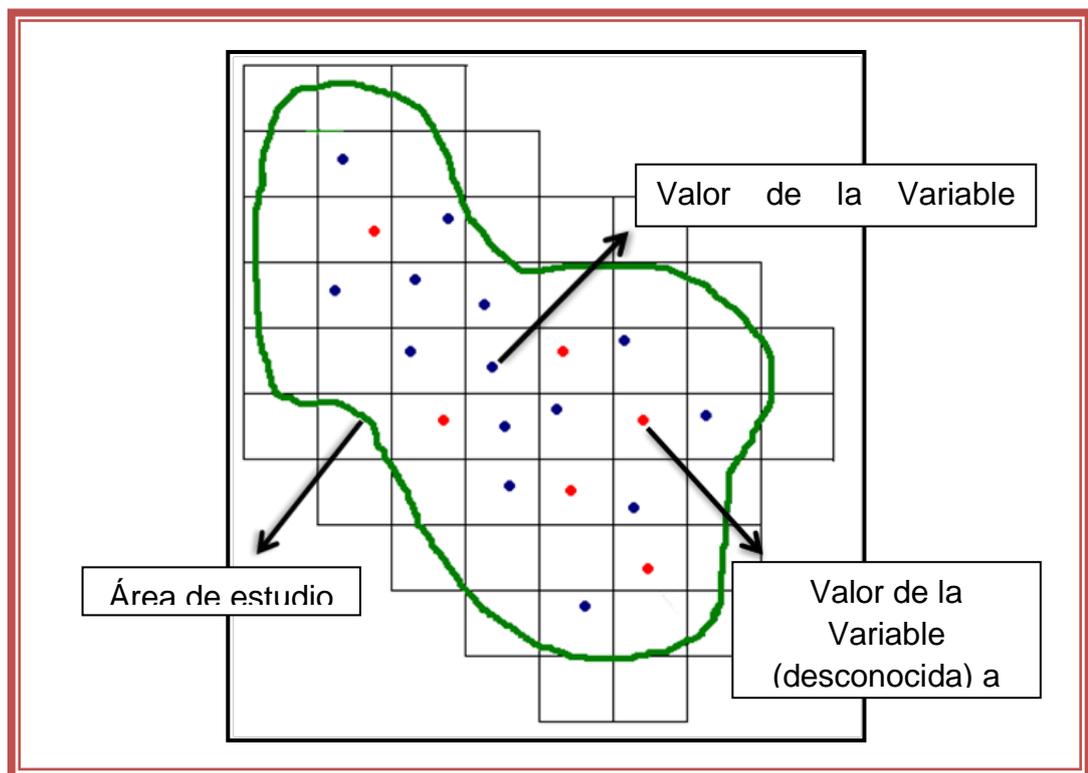


Figura 1.5 Representación de una Predicción Espacial.

Elaborado por: Dennis Lucin

Una distribución de probabilidad no consigue ser fijada con solo unos cuantificaciones; la media y la desviación estándar, no obstante de gran ayuda, no son suficientes, ya que dos variables aleatorias podrían tener la misma media y la misma desviación, pero aun así sus distribuciones pueden

ser muy diferentes. Sin embargo, si conocemos que una variable aleatoria se distribuye de una forma Gaussiana la media y la varianza serán suficientes para definirla completamente. (A. Castaño, F. Vergara, 2004)

1.7.1 KRIGING

El término Kriging es atribuido al geólogo D. G. Krigen creada cuando realizó una estimación de unas minas de oro. El Kriging no es más que una técnica de la geoestadística, que se basa en la interpolación como método matemático para la estimación de valores desconocidos de variables distribuidas espacialmente (la figura 1.5 es un claro ejemplo del kriging).

Este método está relacionado con las siglas B.L.U.E que se refiere al mejor estimador lineal insesgado (Best linear unbiased estimator, por sus siglas en ingles). Está formada por la combinación lineal de datos disponibles (conocidos), la cual demuestran dos características básicas:

- Minimiza la varianza del error (el cuadrado de las desviaciones).
- Para ser insesgado debe cumplir que la suma de los errores sea cercana a 0.

El cálculo de variogramas descrita anteriormente justamente está basado en esta metodología para su cálculo pero con su respectiva ecuación (Ec. 8). El kriging matemáticamente se calcula de la siguiente forma: Sea Z una variable (a estimar) en un punto X_0 , ésta realiza una combinación lineal a partir de las variables vecinas (valor conocido) más cercana al punto de interés X_0 , y realiza el cálculo a través de la ecuación:

$$Z(X_0) = \lambda_1 Z(X_1) + \dots \lambda_n Z(X_n) \quad (9)$$

$$\bar{Z}(X_0) = \sum_i^n \lambda Z(X_i) \quad (10)$$

$$\sum_{i=1}^n \lambda_i = 1 \quad (11)$$

Donde: $Z(X_i)$: Valores en los sitios donde hubo medición.

λ_i : Vector de ponderación que minimiza el error de predicción.

La ecuación 9 y 10 simplemente indica la suma ponderada sobre todos los sectores que conforman la zona de estudio. Mientras que la ecuación 11 muestra la condición que debe cumplir para minimizar el error de predicción resultante, teniendo en cuenta las características geométricas del problema. Estas ponderaciones son fijados a partir de un análisis espacial, establecido en el variograma experimental.

Con respecto a otros métodos el Kriging utiliza un método similar a la interpolación por media móvil ponderada. Existen varios tipos de Kriging, la tabla 1.1 indica estos tipos y sus propiedades.

TIPO DE PREDICTOR	NOMBRE	PROPIEDADES
LINEAL	<ul style="list-style-type: none"> ➤ Simple ➤ Ordinario ➤ Universal 	<ul style="list-style-type: none"> ➤ Son óptimos si hay normalidad multivariada. ➤ Independiente de la distribución son los mejores predictores linealmente insesgados.
NO LINEAL	<ul style="list-style-type: none"> ➤ Indicador ➤ Probabilístico ➤ Log Normal, TransGaussiano ➤ Disyuntivo 	<ul style="list-style-type: none"> ➤ Son predictores óptimos.

Tabla 1.1 Tipos de predictores kriging y sus propiedades.

Fuente: Modificado de Introducción a la Geoestadística, Ramón Giraldo Henao.

1.7.2 KRIGINGSIMPLE

Matemáticamente el Kriging simple es el más sencillo, pero el menos empleado. Se supone que la expectativa del campo aleatorio a ser conocido, se basa en una función de covarianza. En este tipo de kriging se asume que hay una variable regionalizada estacionaria con media m (en todo el área de estudio) y covarianza conocidas. Por supuesto, es poco práctico ya que en general estos dos parámetros son desconocidos y es preciso estimarlos a partir de los datos de la muestra. Para asegurarse de que sea un estimador insesgado debe cumplirse la condición de que $m=0$. Las funciones prácticas para la empleamiento del kriging simple son los siguientes:

- amplia estacionariedad sentido del campo.
- La expectativa es cero en todas partes.
- Función de covarianza conocido.

Para construir el mejor estimador lineal insesgado en este tipo de kriging se la realiza de la siguiente manera: Sea $Z(u)$ la variable de interés medida en el sitio u :

$$Z^*(u) = a + \sum_{i=1}^n \lambda_i * Z(u_i) \quad (12)$$

De las cuales se debe examinar las condiciones de insesgo y de varianza mínima para poder determinar el coeficiente a y los ponderados $\{\lambda_i, i=1...n\}$. Añadiendo la condición de insesgo (valor esperado del error de estimación), condición de varianza mínima (expresada en función de covarianza), es decir los ponderados óptimos, que se la pueden determinar tomando derivadas parciales con respecto a los ponderadores e igualando a cero y aplicando notación matricial se obtiene la siguiente ecuación:

$$Z^*(u) = \sum_{i=1}^n \lambda_i * Z(u_i) + \{1 - \sum_{i=1}^n \lambda_i\} * m \quad (13)$$

Esta ecuación (13) representa a la media con un ponderador que es el complemento de la ponderación acumulada de los datos. Mientras más lejos el sitio u de los datos, menores serán sus ponderadores y mayor será la ponderación de la media. De este modo, la media suple la falta de información contribuida por los datos.

1.7.3 KRIGING ORDINARIO

El kriging ordinario es utilizado en el momento que la variable es estacionaria con covarianza conocida y media desconocida, al desconocer la media de la variable regionalizada permitirá generalizar al kriging a situaciones donde esta media no es constante en el espacio. Aunque el proceso es similar al del kriging simple, no podemos centrar la variable, ya que no conocemos u , así que es necesario trabajar directamente con la variable en estudio Z .

1.7.4 KRIGING UNIVERSAL

Este tipo de kriging considera que la media local no es conocida y varía suavemente en cada vecindad local. La componente de tendencia se modeliza como una combinación lineal de funciones de las coordenadas. Para tratar este tipo de variables es frecuente descomponerla variable como la suma de la tendencia, tratada como una función determinística, más una componente estocástica estacionaria de media cero.

1.7.5 KRIGING RESIDUAL

La técnica kriging residuales empleada bajo las mismas circunstancias del kriging universal, es decir en aquellos casos en que la variable regionalizada

no es estacionaria debido a la presencia de tendencia espacial en el valor promedio de la variable. La hipótesis central del kriging residual consiste en suponer conocida la tendencia $m(x)$, es decir la media. A partir de ella se calculan los residuos con base en los cuales se aplica kriging ordinario. La estimación de la tendencia es generalmente llevada a cabo por medio de mínimos cuadrados. La predicción en un sitio no muestreado es igual a la tendencia estimada más la predicción del error. (Ramón Giraldo Henao.)

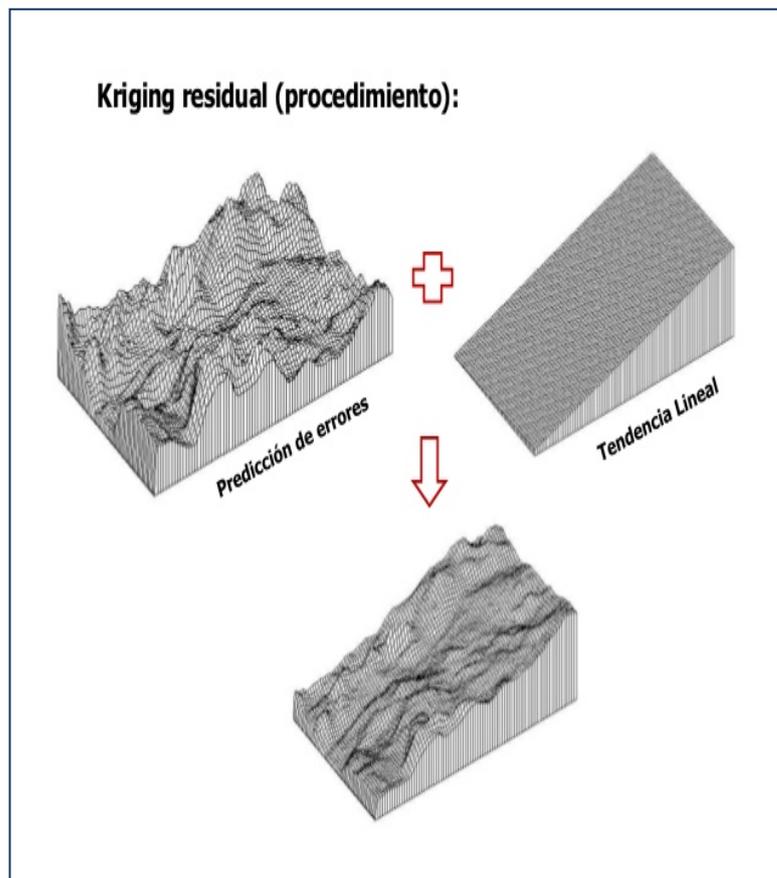


Figura 1.6 Representación del procedimiento kriging residual.

Fuente: Modelamiento de datos climáticos (Geoestadística), Neftalí Chapi S.

La figura 1.6 muestra la representación gráfica del proceso que efectúa el kriging residual donde la superficie interpolada (abajo) es igual a la suma de

la tendencia lineal (arriba derecha) más la predicción de los errores (arriba izquierda).

1.7.6 KRIGING INDICADOR

La predicción de la función indicadora en el sitio x_0 es igual a una combinación lineal de las n funciones indicadoras evaluadas en los sitios de medición. Samper y Carrera (1990) muestran que el kriging indicador es un estimador de la probabilidad acumulada hasta el límite Z definido en la función indicadora.

1.7.7 KRIGING LOG-NORMAL Y MULTI-GAUSSIANO

En estos dos procedimientos se hacen transformaciones de la variable regionalizada con el propósito de normalizar en cada sitio de la región de estudio. Primero consiste en aplicar kriging ordinario a la transformación logarítmica de los datos. Respecto al kriging multi-gaussiano, suponga que se tiene una variable regionalizada $\{Z(x): x \in D\}$ estacionaria. Este procedimiento consiste en hacer una transformación de $Z(x)$ tal que los valores transformados sigan una distribución normal estándar. En ese sentido es una generalización del kriging log-normal. Los pasos del método kriging multi-gaussiano son los siguientes:

- Se encuentra la función de probabilidad acumulada empírica $F_n(Z(x))$.
- Se calculan con base en $F_n(Z(x))$ los “scores” normales estándar, es decir los valores de una distribución de probabilidad normal estándar para los cuales la probabilidad acumulada corresponde a $F_n(Z(x))$.

Capítulo II

2 INTELIGENCIA ARTIFICIAL

2.1 INTRODUCCIÓN

La inteligencia artificial (IA) es un campo muy amplio que comprende temas referentes a la filosofía como la ciencia y la tecnología. Esto implica que han sido muchos los avances realizados para conseguir máquinas capaces de realizar procesamientos cognitivos del ser humano. De hecho es un proyecto ambicioso para la informática que se ha convertido en un tema controversial pero muy prometedor.

Esta unidad corresponderá a los elementos del marco referencial de la inteligencia artificial, partiendo de una idea general del surgimiento de este tema innovador, pero haciendo énfasis en algunas técnicas como Algoritmos Genéticos (AG) que se considerará como metodología dentro de esta tesis. Es necesario e importante la identificación y la comprensión de la complejidad teórica, conceptual e histórica de la inteligencia artificial para el entendimiento del mismo.

2.2 DEFINICIÓN: INTELIGENCIA ARTIFICIAL

Para poder definir conceptualmente lo que es Inteligencia Artificial es necesario conocer el significado de estas dos palabras para de esta forma establecer el concepto general. Existen varios conceptos de la palabra "Inteligencia", citados por varios autores, de las cuales algunas son las siguientes:

- La habilidad para pensar y para desarrollar el pensamiento abstracto, como capacidad de aprendizaje, como manipulación, procesamiento, representación de símbolos, capacidad para adaptarse a situaciones nuevas, o para solucionar problemas (Mayer, 1983)

- Inteligencia, es la fuerza intelectual, la capacidad de conocer o de entender. El valor en que un individuo logra resolver favorablemente una nueva situación o un problema. La inteligencia está establecida en el nivel de conocimientos individuales y en la habilidad de manipular y reformular apropiadamente los conocimientos en base a los datos que se proporcionan como requerimientos para resolver algún problema o situación. (Arauz, 1998)

De la misma manera la palabra Artificial citados por algunos autores definen a esta como:

- “Lo artificial [...] radica en el resultado de los esfuerzos premeditados de lograr los mismos resultados de la naturaleza a través de estrategias diferentes de aquellas que la naturaleza sigue, como la tecnológica” (Massimo Negrotti).
- Artificial, es aquel cuyo producto origen es no natural, sino que fue hecho por la mano o arte del hombre. (Zampayo, 2004)

Conociendo el significado de ambas palabras podemos definir a la Inteligencia Artificial como una rama de la computación que, a través de programas y algoritmos de computador se vincula un fenómeno natural con la afinidad artificial. Podemos citar varios conceptos para aclarar el tema en sí.

- La inteligencia artificial (IA) es una rama de la informática que intenta reproducir las funciones cognitivas humanas como el razonamiento, la memoria, el juicio o la decisión y, después, confiar una parte de esas facultades, que consideramos signos de inteligencia, a los ordenadores. (Bourcier, 2003)

- La inteligencia artificial (IA) es una de las áreas más fascinantes y con más retos de las ciencias de la Computación ya que ha tomado a la inteligencia como la característica universalmente aceptada para diferenciar a los humanos de otras criaturas ya sean vivas o inanimadas, para construir programas o computadoras inteligentes. (Gutiérrez, 2006, p.11)

2.3 PANORAMA HISTÓRICO

Es necesario conocer el surgimiento e historia de la IA destacando los acontecimientos y logros más relevantes. La IA ha tenido más relevancia o mayor desarrollo en las últimas décadas, pero se describirá los sucesos que hicieron que esta ingresara en nuevas áreas de investigación.

De esta forma ayudaron al desarrollo en tareas de resolución de problemas en campos especializados, como en la realización de análisis químicos, en el campo de la ingeniería (diseño, detección de fallos, planificación de manufacturación, etc.), en el análisis científico, en la medicina, en el análisis financiero, etc. según (Huerta, 2009, p. 18). A Continuación se detalla la historia de la Inteligencia Artificial.

Año	Descripción
1642	Blaise Pascal, perfecciona la pascalina, una máquina capaz de realizar sencillas operaciones como sumar y restar. Es considerada como la primera calculadora automática mundial.
1694	Gottfried Wilhem Leibniz, perfecciona una computadora con el nombre de "Leibniz", la cual puede realizar multiplicaciones, por medio de un algoritmo que realiza sumas de manera repetitiva, dicho algoritmo aun es utilizado en algunas computadoras.

1832	Charles Babbage desarrolla la primera computadora mundial que puede ser programada para resolver una amplia variedad de problemas lógicos y computacionales
1943	La base de la Inteligencia Artificial es asentada, gracias a Warren Mc Culloch y Walter Potts, quienes propusieron un modelo de neurona de cerebro humano y animal, esto proporciono una representación simbólica de la actividad cerebral.
1946	Es desarrollada la primera computadora completamente electrónica y digital programable, llama ENIAC. Fue desarrollada por Jonh Prespert Eckert y Jonh W. Mauchley.
1950	Alan Turing describe los medios para determinar cuándo una máquina es inteligente, a lo que se llamo “prueba de Turing”.
1955-1956	El primer programa de IA, fue escrito por Allen Newell, Herbert Simon y Shaw JC. Demostró teoremas usando una combinación de búsqueda, el comportamiento orientado a objetos, y la aplicación de las normas.
1958	John McCarthy introduce LISP, un primer lenguaje de Inteligencia Artificial.
1960	Se inicia el estudio de las estructuras sintácticas de la computadora y sus relaciones con las estructuras sintácticas del lenguaje humano.
1965	Edward Feigenbaum y Robert K. Lindsay construyen en Stanford DENDRAL, el primer sistema experto. Mapeaba la estructura de los productos químicos orgánicos complejos de datos de un espectrómetro de masas.
1975	La medicina se ha convertido en un área importante para las aplicaciones de la investigación en Inteligencia Artificial.

Inicios de 1980	Una sucesión de sistemas expertos fueron construidos y puestos en uso por las empresas.
1990	Se han creado empresas sobre Inteligencia Artificial y el modelo conexionista empieza a cobrar mayor fuerza como paradigma frente al modelo de procesamiento simbólico, a partir del Primer Congreso Internacional sobre Redes Neuronales.
1997	El programa de ajedrez Profundo Azul gana al actual campeón mundial de ajedrez, Garry Kasparov, en un partido muy reñido.
1999	Sony Corporation presentó el AIBO, un perro robot mascota que entiende comandos de voz, aprende y madura. AIBO es un acrónimo de Inteligencia Artificial robot, Aibo.
2000	Cynthia Breazeal en el MIT publica su disertación sobre las máquinas Sociable, describiendo Kismet, un robot con una cara que expresa emociones.
2002	iRobot, fundada por investigadores de la MIT Laboratorio de Inteligencia Artificial, presenta Roomba, una aspiradora robot de limpieza. Para el año 2006, dos millones se habían vendido.
2004	En julio, investigadores de la Universidad Californiana de Pasadena, en Estados Unidos, lograron registrar las señales que emiten las neuronas relacionadas con la planificación de los movimientos del cuerpo en monos, y decodificarlas usando un ordenador.
2009	Desarrollaron un sistema que permite registrar el estado emocional de niños autistas. Nilanjan Sarkar (uno de sus creadores).

Tabla 2.1 Historia de la Inteligencia Artificial

Fuente: Berenice Hernández Miranda Jorge, Alfonso Haces Álvarez

2.4 TÉCNICAS DE LA INTELIGENCIA ARTIFICIAL

Para la solución de problemas o búsquedas de soluciones, la Inteligencia Artificial se basa en diferentes técnicas como herramientas que son utilizadas desde el punto de vista ingenieril, cada uno aporta con una metodología según al área a aplicarse. En la siguiente figura 2.1 se señala algunas de las técnicas de la Inteligencia Artificial.



Figura 2.1 Técnicas de la Inteligencia Artificial.

Elaborado por: Dennis Lucin

De las técnicas señaladas anteriormente, se detallará brevemente varias de ellas y a profundidad lo que corresponde a los Algoritmos Genéticos que es la metodología que se empleará en los capítulos posteriores.

2.4.1 ROBÓTICA

Dentro del campo de la robótica la Inteligencia artificial ha demostrado que logra resolver complejos problemas de carácter masivo, por lo cual ha desarrollado varios aspectos en sistemas tales que:

- Permiten al usuario preguntar a una base de datos en cualquier lenguaje, que sea mejor que un lenguaje de programación.
- Reconocen objetos de una escena por medio de aparatos de visión.
- Generar palabras reconocibles como humana desde textos computarizados.
- Reconocen e interpretan un pequeño vocabulario de palabras humanas.
- Resuelven problemas en una variedad de campos usando conocimientos expertos codificados.

2.4.2 REDES NEURONALES

Una red neuronal es un sistema de procesadores paralelos conectados entre sí en forma de grafo dirigido. Esquemáticamente cada elemento de procesamiento (neuronas) de la red se representa como un nodo. Estas conexiones establecen una estructura jerárquica que tratando de emular la fisiología del cerebro, busca nuevos modelos de procesamiento para solucionar problemas concretos del mundo real (figura 2.2). Lo trascendental en el desarrollo de la técnica de las red neuronal es su útil comportamiento al aprender, reconocer y aplicar relaciones entre objetos y tramas de objetos propios del mundo real. (Freman y Skapura)

Existen Varios principios como parte fundamental para el desarrollo de problemas complejos en la que se basa una red neuronal, las más relevantes son:

- Aprendizaje adaptativo.
- Autoorganización.
- Tolerancia a fallos.
- Operación en tiempo real.
- Fácil inserción en la tecnología existente.

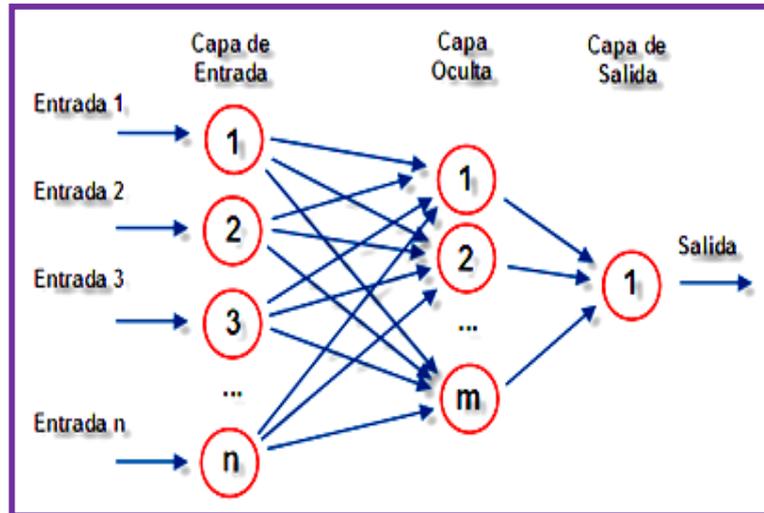


Figura 2.2 Esquema de una Red Neuronal

2.4.3 PROCESAMIENTO DEL LENGUAJE NATURAL

Procesamiento del Lenguaje Natural (PLN) es una herramienta de la inteligencia artificial y rama de la ingeniería lingüística computacional. El lenguaje natural es un fenómeno muy complejo, pero se ha demostrado que las expresiones del lenguaje humano están organizadas a través de un conjunto de reglas. De esta manera el lenguaje humano ha sido vinculado al lenguaje computacional, el fin de esta técnica es lograr una comunicación humano-máquina equivalente a la comunicación humano-humano

2.4.4 BÚSQUEDA DE SOLUCIONES

Las búsquedas proporcionan una forma de resolver los problemas en los que no se dispone de un método más directo como una estructura con técnicas directas existentes según (Arauz, 1998, p. 4).

Las heurísticas son criterios, métodos o principios para decidir cuál de entre varias acciones promete ser la mejor para alcanzar una determinada meta.

2.5 COMPUTACIÓN EVOLUTIVA - ALGORITMOS GENÉTICOS

Los Algoritmos Genéticos de ahora en adelante AG, es una técnica de la Inteligencia Artificial que ha tenido mucha relevancia en la computación evolutiva como parte fundamental, que constituyen un nuevo enfoque a la solución de innumerables problemas para la optimización y búsqueda. Referente a la historia se puede identificar tres aspectos principales sobre la computación evolutiva:

- La programación evolutiva que fue establecido por Fogel, Owens, y Walsh en 1966.
- Las estrategias evolutivas por Rechenberg en 1973.
- Los algoritmos Genéticos establecido por Holland en 1975.

La computación evolutiva representa informáticamente a la evolución natural con la ejecución de herramientas computacionales para la resolución de problemas. Comenzando la década de los 60, fueron presentados y sugeridos varios modelos de computación evolutiva, que colectivamente se les atribuyo como algoritmos evolutivos. Estos algoritmos han sido utilizados para una extensa diversidad y complejidad de problemas hallados tanto en la industria y el comercio, como en la investigación científica.

Antes de referirnos detalladamente a los AG, revisaremos brevemente algunos aspectos en la que está basada esta herramienta como la teoría de la evolución y las leyes genéticas.

2.5.1 TEORÍA DE LA EVOLUCIÓN

La teoría de la evolución fue establecida por el naturalista Charles R. Darwin (1809-1882), en la expedición del Beagle por América del sur y el Pacífico en el año 1831. En dicha expedición realizó una variedad de observaciones para lo cual obtuvo una serie de información copilada dando como resultado de su trabajo ciertos postulados que se presenta a continuación:

1. Las formas de vida no son estáticas sino que evolucionan y las especies cambian continuamente, unas se originan y otros se extinguen.
2. El proceso de la evolución es gradual, lento y continuo, sin saltos discontinuos o cambios súbitos.
3. Los organismos parecidos se hallan emparentados y descienden de un antepasado común. Todos los organismos vivientes pueden remontarse a un origen único de la vida.
4. La selección natural es la llave, en dos fases, que explica todo el sistema. la 1ra fase es la producción de variabilidad: la generación de modificaciones espontáneas en los individuos. la 2da fase, la selección a través de la supervivencia en la lucha por la vida: los individuos mejor dotados, los que han nacido con modificaciones espontáneas favorables para hacer frente al medio ambiente van a tener más posibilidades de sobrevivir, de reproducirse y de dejar descendencia con estas ventajas.

2.5.2 LEYES DE LA GENÉTICA

Las leyes de la genética se la atribuyen a Gregor Mendel gracias a sus experimentos realizados durante aproximadamente 8 años. La observación de los resultados obtenidos llevó a Mendel a formular algunas suposiciones que explicaran los resultados obtenidos:

1. “En cada organismo hay un par de factores que controlan la manifestación de una cualidad particular”. Es decir que en sus observaciones notó que existen factores que determinan la característica de cada generación a través de los rasgos parenteral
2. “Si un organismo tiene dos factores antagónicos para una característica, uno de ellos puede expresarse con exclusión total del otro”.

Mendel supuso que el gen con la característica observada era más “poderoso” que el determinante del rasgo excluido. De esta forma denominó **gen dominante** al que causa el efecto a su generación; y **gen recesivo** al que no se manifiesta en presencia del dominante.

Con estas observaciones Mendel estableció dos leyes las cuales son:

Ley de la Segregación: Los determinantes hereditarios son de naturaleza particulada. Estos determinantes son denominados en la actualidad genes. En los individuos diploides cada individuo posee un par de estos determinantes o genes en cada célula para cada característica estudiada. Todos los descendientes de un cruzamiento de dos líneas puras (Generación¹) tienen un alelo para el fenotipo dominante y uno para el fenotipo recesivo. Estos dos alelos forman el par de genes. Un miembro del par de genes segrega en cada gameto, de manera que cada gameto lleva solamente un miembro del par de genes (figura 2.3). (Mendelian Genetics)

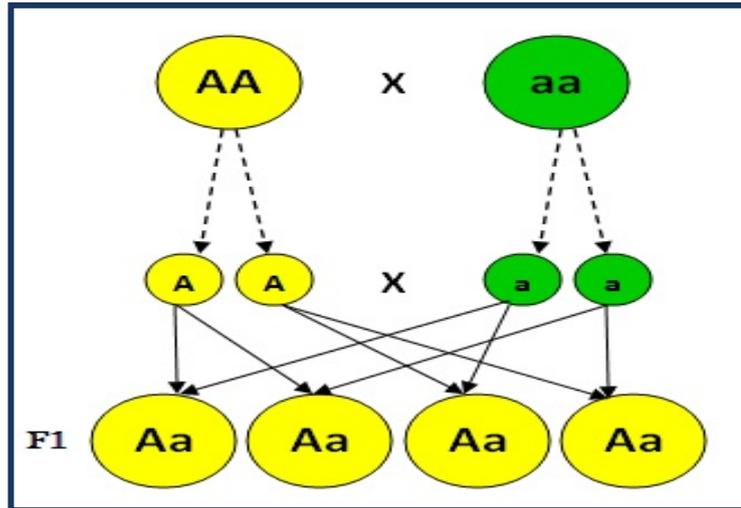


Figura 2.3 Ley de la Segregación

Fuente: <http://www.saberpractico.com/>

Ley de la Segregación Independiente: Esta segunda ley establece que durante la formación de los gametos cada alelo de un par se separa del otro miembro para determinar la constitución genética del gameto filial como se muestra en la figura 2.4. Es muy habitual representar las posibilidades de hibridación mediante un cuadro de Punnett.

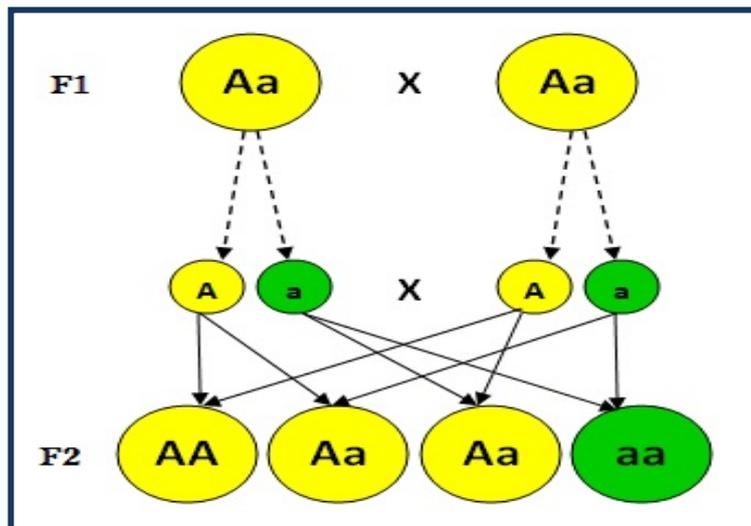


Figura 2.4 Ley de la Segregación Independiente

Fuente: <http://www.saberpractico.com/>

Una vez detallado los aspectos en la que se basa los AG entraremos a definir todo relacionado a los Algoritmos Genéticos, que es la parte importante dentro de este proyecto.

2.5.3 ALGORITMOS GENÉTICOS

Como ya habíamos mencionado anteriormente quien estableció las bases para su posterior desarrollo fue John Holland en 1975. Los algoritmos genéticos es un acierto más de la Inteligencia Artificial en su afán de imitarlos comportamientos biológicos, mediante la computación con el beneficio de los avances científicos que ello implica. Estos algoritmos están basados en la mecánica de la selección natural y leyes de la genética que se puntualizó anteriormente.

2.5.4 DEFINICIÓN DE AG

John Cosa propone una definición bastante explícita y completa sobre el significado de Algoritmos Genéticos, la cual nos indica que un AGes un algoritmo matemático altamente paralelo que transforma un conjunto de objetos matemáticos individuales con respecto al tiempo usando operaciones modeladas de acuerdo al principio Darwiniano de reproducción y supervivencia del más apto, y tras haberse presentado de forma natural una serie de operaciones genéticas de entre las que destaca la recombinación. Cada uno de estos objetos matemáticos suele ser una cadena de caracteres (letras o números) de longitud fija que se ajusta al modelo de las cadenas de cromosomas, y se les asocia con una cierta función matemática que refleja su aptitud.

2.5.5 ESQUEMA DE UN AG

El esquema más sencillo o básico es la que se describe a continuación en una serie de procesos a ejecutar:

- Evaluar la puntuación de cada uno de los cromosomas generados.
- Permitir la reproducción de los cromosomas siendo los más aptos (mejores características) los que tengan más probabilidad de reproducirse.
- Con cierta probabilidad de mutación, mutar un gen del nuevo individuo generado.
- Organizar la nueva población.
- Repetir el proceso con la nueva generación (Hasta conseguir el individuo solución al problema)

En la figura 2.5 se presenta el diagrama de flujo con los segmentos más significativos de un algoritmo genético básico: selección, combinación y sustitución. Aparte de estos procesos, que conciernen al eje principal y propio de los algoritmos genéticos, de igual importancia existen otros, que sirven de soporte a los anteriores: descripción del problema a resolver, codificación de sus soluciones e inicialización. Selección y codificación son los puntos críticos de un algoritmo genético.

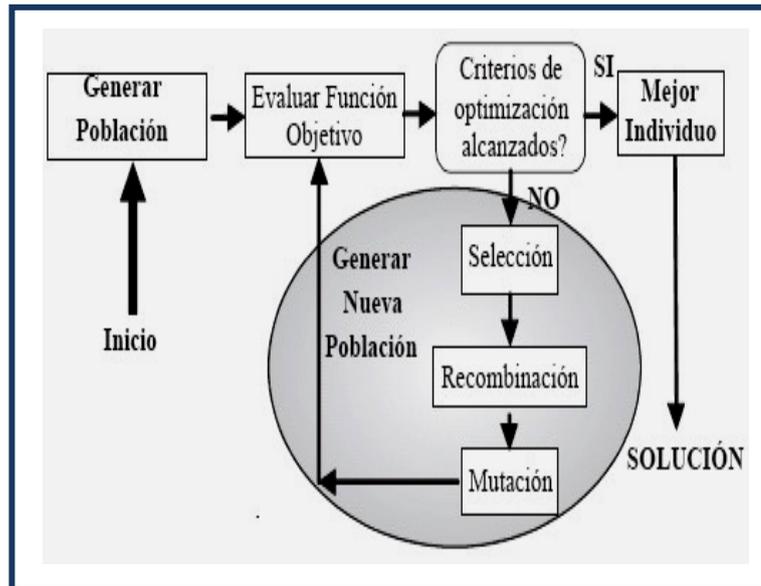


Figura 2.4Diagrama de un algoritmo evolutivo

Fuente: Yllada García, Ruth A. Maneiro Malavé, Ninoska

2.5.6 PARÁMETROS DE LOS ALGORITMOS GENÉTICOS.

Para la comprensión y desarrollo de un algoritmo genético es necesario tomar en cuenta una serie de parámetros a considerar tales como:

2.5.6.1 TAMAÑO DE LA POBLACIÓN

La población inicial por lo general se elige formando cadenas al azar. Esta medida nos revela el número de cromosomas que poseemos en nuestra población para una reproducción determinada. En caso de que éste régimen sea insuficiente, el algoritmo genético poseerá escasas posibilidades de realizar o producir generaciones con lo que se efectuaría una búsqueda de soluciones insuficientes y poco óptimas. Por otra parte si la población es descomunal, el algoritmo genético ejecutaría lentamente. Existen estudios que muestran que hay un límite a partir del cual es ineficiente engrandecer el tamaño de la población, ya que no se lograría una mayor rapidez en la resolución del problema.

2.5.6.2 PROBABILIDAD DE CRUCE

El Algoritmo Genético Canónico o básico, utiliza el cruce fundamentado en un punto, donde los dos especímenes o individuo seleccionados para realizar el título de padres, son recombinados mediante la selección de un punto de corte, ésta a su vez intercambia las secciones que se encuentran a la derecha de dicho punto.

Dicho de otra manera este parámetro indica la reiteración con la que se generan los cruces entre los cromosomas padres, es decir, que haya posibilidad de reproducción entre ellos.

2.5.6.3 POSIBILIDAD DE MUTACIÓN

La mutación se considera como un indicador básico, que provee un pequeño componente de aleatoriedad en el entorno de los individuos de la población. Esto indica la frecuencia con la que los genes de un cromosoma le suceden un cambio radical o mutación.

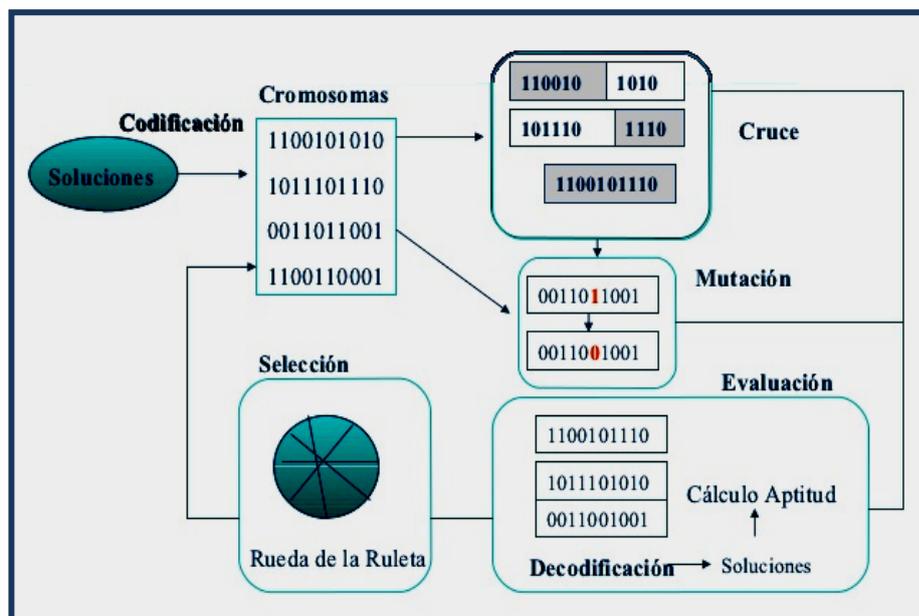


Figura 2.5 Estructura de un algoritmo genético simple.

Fuente: Inteligencia Artificial, Milla José, Pérez Jemima

2.5.7 APLICACIONES DE LOS ALGORITMOS GENÉTICOS

Los AG han servido para que los estudios efectuados en torno a él, se hayan empleado a una variedad de problemas y tipos en ingeniería, y en la ciencia en general. Algunas aplicaciones en las que los AG se ha destacado son:

- **Optimización:** Los Algoritmos Genéticos se han usado en cuantiosos trabajos de optimización, abarcando también la optimización numérica, y los problemas de optimización combinatoria.
- **Programación automática:** Los AG se han utilizado para desarrollar programas exclusivamente para cumplir tareas específicas, y para diseñar otras estructuras computacionales tales como el autómata celular, y las redes de clasificación.
- **Aprendizaje máquina:** Los algoritmos genéticos se han utilizado también en muchas de estas aplicaciones, tales como la predicción del tiempo o la estructura de una proteína. Han servido asimismo para desarrollar determinados aspectos neuronales, las reglas para sistemas de clasificación de aprendizaje o sistemas de producción simbólica, y los sensores para robots.
- **Sistemas inmunes:** A la hora de modelar varios aspectos de los sistemas inmunes naturales, incluyendo la mutación somática durante la vida de un individuo y el descubrimiento de familias de genes múltiples en tiempo evolutivo, ha resultado útil el empleo de esta técnica.
- **Genética de poblaciones:** En el estudio de preguntas del tipo “¿Bajo qué condiciones será viable evolutivamente un gen para la recombinación?”

- **Evolución y aprendizaje:** Los Algoritmos Genéticos se han utilizado en el estudio de las relaciones entre el aprendizaje individual y la evolución de la especie. Sistemas sociales: En el estudio de aspectos evolutivos de los sistemas sociales, tales como la evolución del comportamiento social en colonias de insectos, y la evolución de la cooperación y la comunicación en sistemas multi-agentes. sistemas particulares de aprendizaje, como pueda ser el de los pesos en una red. (Mitchell M. 1998)

2.5.8 CODIFICACIÓN DE PROBLEMAS

Puede ser mostrada cualquier solución a un problema proporcionando ciertos valores a una serie de parámetros (genes). Las soluciones codificadas en un cromosoma luchan para ver cuál establece la mejor solución al problema. Consecuentemente, un algoritmo genético radica en encontrar de qué parámetros depende el problema, codificarlos en un cromosoma, y utilizar las reglas de la evolución. Por lo tanto hay que estudiar la codificación más óptima según el caso o problema que se esté estudiando.

2.5.8.1 CODIFICACIÓN BINARIA

Esta forma de codificación es la más empleada, ya que mediante la combinación de números binarios, puede abarcar muchos cromosomas en una cadena de bits (1 y 0) inclusive con un escaso número de genes. La figura 2.6 presenta un modelo de un individuo binario que codifica 3 parámetros mientras que la figura 2.5 es un claro ejemplo del proceso de ejecución de un AG con codificación binaria.

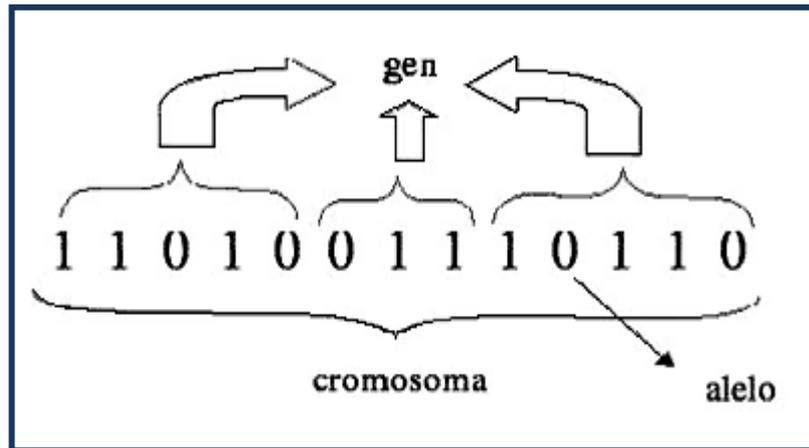


Figura 2.6 Individuo Genético Binario.

Fuente: M. Gestal

2.5.8.2 CODIFICACIÓN NUMÉRICA

En este tipo de codificación se utilizan series de números que simbolizan un número en una sucesión. Se utiliza en problemas en los que hay que organizar ciertos parámetros, donde resulta muy ventajoso.

2.5.8.3 CODIFICACIÓN POR VALOR DIRECTO

Este tipo de codificación será el utilizado en caso de resolución de problemas en el que se requiera del uso de valores de cifrado complicado como podría ser en el uso de números reales, cuya codificación con números binarios sería muy complejo. (Parra Truyol, Antonio)

2.5.8.4 CODIFICACIÓN EN ÁRBOL

Es usado especialmente en el desarrollo de proyectos para programación genética. Cada cromosoma corresponderá en este caso un árbol con ciertos elementos.

En esta técnica, los cambios al azar logran formarse cambiando el operador, trastornando el valor de un cierto nodo del árbol o simplemente suplantando un subárbol por otro.

Capítulo III

3 LOS VARIOGRAMAS

3.1 INTRODUCCIÓN

Los valores de una variable regionalizada descrita en el capítulo 1 no son independientes, refiriéndonos en la comprensión de que un valor observado en una zona, suministra información sobre los valores de las zonas cercanas. En la deducción probabilística de la variable regionalizada, esta noción intuitiva de dependencia está puntualizada por la distribución espacial de la función aleatoria, que ajusta la manera como se relacionan los valores estudiados en distintos sitios por una distribución de probabilidad multivariable.

La interpolación espacial es un problema clásico de estimación de una función en un punto de un plano a partir de valores conocidos en un cierto número de puntos regionales, que es la metodología usada para el modelamiento geoestadístico basada en celdas. El problema consiste en determinar los valores de las variables en sitios que se desconocen su valor en función de qué parámetro (ponderación) se efectuará el cálculo para la interpolación espacial. Para esto existen muchas maneras o metodologías de resolver dicho problema y una de ellas es el kriging tal como se lo mencionó en el capítulo 1.

Dentro de las técnicas que se emplean para describir la relación espacial, el variograma es la más utilizada debido a que comprende dicha relación y la expresa gráficamente para su análisis. El kriging utiliza el variograma como método matemático para la estimación de valores desconocidos de variables distribuidas espacialmente.

En este capítulo, abordamos la etapa de los variogramas, que consiste en el detalle y descripción de lo relacionado con el cálculo del mismo, parámetros de consideración, elementos que la componen, análisis variográfico, anisotropías y su posterior modelamiento entre otros aspectos. Es de vital importancia el entendimiento de los variogramas, es decir, el papel que

desempeña en el modelamiento geoestadístico y cada aspecto que lo componen para la posterior programación (capítulo 4) que efectuará el cálculo de variogramas y su respectiva aplicación de los algoritmos genéticos que se empleará como medida de optimización del mismo.

3.2 DEFINICIÓN DE VARIOGRAMAS

El variograma es la técnica geoestadística más comúnmente usada para capturar la relación espacial. Es un instrumento que permite estudiar el comportamiento espacial de una variable (propiedad) sobre una región dada y muestra como dos valores en el espacio se ponen en correlación. Matemáticamente se la define como la mitad de la varianza de la diferencia de dos valores situados a una distancia h . Esta ecuación está representada en la sección 1.6.2 del capítulo 1.

$$\bar{\gamma}(h) = \frac{\sum (Z(x+h) - Z(x))^2}{2h}$$

El variograma es un estimador de la varianza poblacional, así que debe tener una preferencia de estacionaridad y es un sustento para las técnicas del Kriging ya que faculta representar cuantitativamente la variación de un fenómeno regionalizado en el espacio. El variograma está ligado con la dirección y la distancia (h).

3.3 ELEMENTOS DE UN VARIOGRAMA

El variograma se ve restringido porque es un estadístico de dos puntos y asimismo porque es extremadamente sensible a valores extremos. Cada elemento está identificado y representado, es así que se ha designado

nombres a dichas características del variograma. Tradicionalmente se ha identificado tres elementos más representativos.

El variograma está formado por los siguientes elementos: La **meseta** es el elemento en la cual la distribución es invariable, es decir permanece constante, en un valor que corresponde a la varianza. El **efecto pepita** (nugget) constituye una diferencia derivado de la existencia de estructuras geológicas o errores en el cálculo y la distancia a la cual el variograma alcanza la meseta se denomina **alcance**.

La figura 3.1 muestra los elementos referentes a un variograma, donde el eje x constituye la distancia, denotada con la letra h, el eje y contiene los valores del variograma en función de la distancia, "a" es una distancia dada cualquiera, C_0 representa el efecto pepita, además se identifica la meseta o sill y el alcance o también llamado rango.

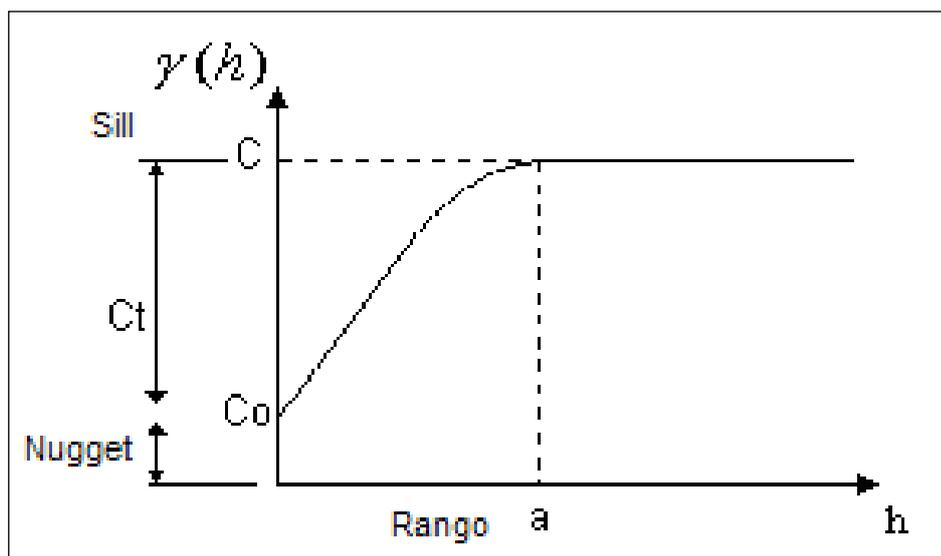


Figura 3.1 Elementos del variograma

Fuente: Evelyn Véliz

3.4 CALCULO DE VARIOGRAMAS

El variograma es una función que se calcula mediante el análisis de los datos distribuidos en el área de estudio. Para el cálculo de variogramas es necesario definir previamente ciertos parámetros con el fin de restringir la influencia de los datos vecinos y obtener una mejor aproximación.

3.4.1 DIRECCIONES

Este parámetro se refiere a la orientación a la cual se realizará la correlación, es decir, la dirección, medida en el plano vertical del azimut, en la que se calcula el variograma. Las direcciones comúnmente utilizadas para el análisis son 0° , 45° , 90° y 135° , estas direcciones tienen como propósito incluir una cantidad de pares de datos que logre capturar el variograma estimado en varias orientaciones, la dirección 0° corresponde al Este y la de 90° al norte.

3.4.2 VALOR DE LAG

Representa los múltiplos de una distancia que indica la separación de los datos con la que se efectuará el variograma experimental, es decir el espaciamiento con lo que se fijará realizar el cálculo.

3.4.3 NÚMERO DE PARES.

Se debe procurar asegurar un número suficiente de pares de datos con el objetivo de restringir y delimitar la distancia máxima a la cual calcularemos el variograma, en primera instancia se podría pensar que la asignación de esta distancia es arbitraria y en cuyo caso sería extremadamente difícil de asignar, pero para evitar esto se ha desarrollado una regla que facilita esta

asignación. Esta norma consiste en establecer al valor de la distancia de “lag” como la mitad de la distancia máxima de separación entre dos puntos de muestras. (Kelkar Y Pérez, 2002)

3.4.4 TOLERANCIA DE LAG.

La tolerancia de lag también es usada para establecer un número adecuado de pares de datos. En un estudio real (zona de estudio) los datos no están distribuidos regularmente lo que conlleva a perder gran cantidad de datos para su análisis. De manera tal que se incluirá muestras que se encuentre entre un porcentaje mayor al lag y menor a la misma.

3.4.5 TOLERANCIA ANGULAR

Ángulo dentro del que se consideran válidos los datos para el cálculo de la diferencia cuadrática, en el mismo plano vertical en que se definió la inclinación o dirección. De la misma forma que en el parámetro anterior la tolerancia angular abarca los datos que se encuentre entre un valor mayor y menor a la dirección

3.4.6 ANCHO DE BANDA

El ancho de banda delimita el rango máximo de influencia de los pares, al ancho de banda se lo puede definir como la distancia perpendicular a la cual se calcula el variograma. Mediante este parámetro y la tolerancia angular quedan definidos el largo y ancho máximos del área que se usarán para los pares seleccionados. (Kelkar Y Pérez, 2002)

3.5 CONSIDERACIONES ASOCIADAS AL CÁLCULO DE VARIOGRAMAS

El variogramas experimental se ve ligado a parámetros que se deben considerar para el entendimiento de los resultados obtenidos. Los aspectos que se necesitan considerar son:

3.5.1 INESTABILIDAD

Si se analiza la ecuación del variograma podemos observar claramente que abarca un factor que representa la diferencia de dos pares de datos al cuadrado, por lo que cuando la diferencia entre los datos es grande esta se ve aumentada ya que está elevada al cuadrado, es decir que si la diferencia entre los pares es demasiado grande el valor del variograma se va a ver afectado creando inestabilidad que puede impedir la captura de la estructura espacial de la información subyacente y que también crea fluctuaciones a medida que la distancia de lag aumenta por lo que se debe procurar disminuir este efecto para poder observar un diagrama interpretable. (Kelkar Y Pérez, 2002)

3.5.2 INFLUENCIA DE LOS VALORES ATÍPICOS

Los valores atípicos son muestras de datos que incurren fuera de lo "normal", los puntos de datos que quedan fuera de más o menos tres desviaciones del promedio de la norma se puede considerar valores atípicos.

Al igual que con el análisis de la inestabilidad los valores atípicos pueden aumentar el efecto que la diferencia cuadrada tiene en el variograma, solamente si la diferencia son varias órdenes de magnitud la diferencia cuadrada se vuelve lo suficientemente grande como para alterar el valor del variograma creando como ya se explicó problemas al identificar la estructura

espacial, es por esto que se deben evitar valores extremos al momento de estimar el variograma. (Kelkar Y Pérez, 2002)

3.6 MODELOS TEÓRICOS DEL VARIOGRAMA

Los modelos teóricos se ajustan a un modelo de variograma experimental y son precisos de explicar los numerosos comportamientos que pueden tener las variables en el espacio. De acuerdo a esto, el variograma requiere ser ajustado a un modelo ya que al ser un cálculo empírico es inconcluso. El propósito de representar el variograma a un modelo, es para definir las características notables que se aprecian en un variograma estimado, empleando modelos preestablecidos a partir de una ecuación que integre los requerimientos necesarios.

Las dos características más importantes en el modelado de Variogramas son:

- 1.- Su comportamiento en el origen.
- 2.- La presencia o ausencia de meseta.

Es de suma importancia modelar los variogramas, ya que los variogramas experimentales solo nos permiten conocer los valores a la distancia de la información muestreada disponible, sin embargo el propósito es estimar estos valores en los lugares no muestreados. Cuando se estime un variograma, no es necesario que las distancias entre las ubicaciones muestreadas y las ubicaciones no muestreadas coincidan con las distancias a las que se estimó el variograma. De hecho, las distancias raramente coinciden con las distancias de la variograma estimado, lo que significa que se tiene que determinar el valor de los variogramas a distancias entre las distancias del lag a la que hemos estimado el variograma. Una opción es simplemente interpolar entre los valores estimados para conocer el valor del variograma a cualquier distancia del lag, aunque, esta aproximación simple

puede crear problemas en el procedimiento de estimación. Por lo que, se tiene que limitar el estudio a ciertos tipos de modelos que pueden ser utilizados para modelar el variograma. (Kelkar Y Pérez, 2002)

3.6.1 EFECTO PEPITA

El modelo Efecto Pepita representa un fenómeno completamente aleatorio, en el cual existe un desfase en la correlación espacial de las muestras independiente en la distancia igual a cero. Su estructura matemática representa una función constante, por definición el variograma es cero a una distancia lag igual a cero, más allá de este punto se mantiene constante y estable, presenta una discontinuidad en el origen y se manifiesta linealmente.

La función matemática para variogramas con efecto pepita es:

$$\gamma(h) = \begin{cases} C \\ 0 \end{cases} \quad (14)$$

3.6.2 MODELO ESFÉRICO

Por lo general, de todos los modelos existentes es el más utilizado para el análisis geoestadístico, es decir el modelo que más se ajusta al variograma experimental ya que es una expresión polinomial simple, en la figura 3.2 se puede distinguir un incremento casi lineal y después a cierta distancia finita del origen logra una equilibrio, la meseta. Está expresado mediante la función:

$$\gamma(h) = C \left[\left(\frac{3h}{2a} \right) - \left(\frac{h^3}{2a^3} \right) \right] \quad \text{para } h \leq a \quad (15)$$

Dónde:

a: Alcance

h: Distancia

C= meseta

Una vez alcanzada el sill o meseta, es decir que se estabiliza, el modelo tiende a comportarse con la siguiente ecuación:

$$\gamma(h) = C \quad \text{para } h > a \quad (16)$$

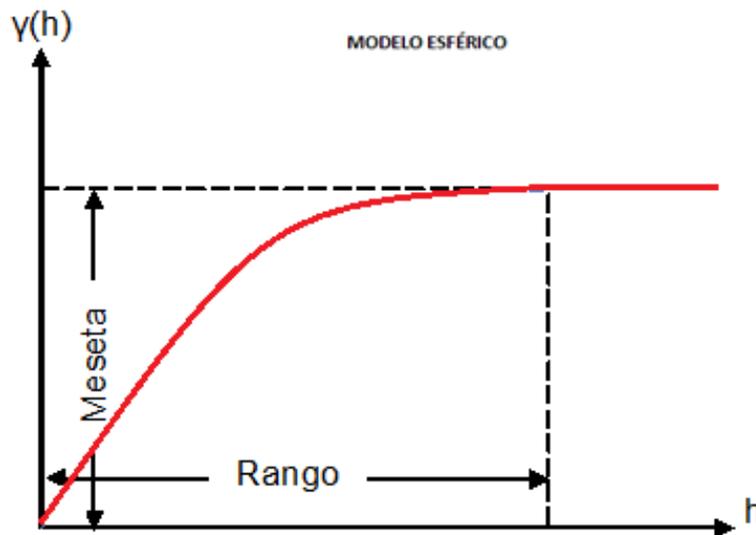


Figura 3.2 Modelo Esférico
Elaborado por: Dennis Lucin

3.6.3 MODELO EXPONENCIAL

En este modelo podemos apreciar que en la figura 3.3 incrementa inicialmente más rápido y después se estabiliza de forma asintótica. Matemáticamente la meseta no se alcanza a una distancia finita, debido a esto se usa con fines prácticos el “alcance efectivo” o “alcance práctico”, se considera que el rango o alcance a , es la distancia para la cual el valor del variograma es el 95% de la meseta, la función que la representa es:

$$\gamma(h) = C \left[1 - \exp\left(-\frac{3h}{a}\right) \right] \quad (17)$$

Dónde:

a: Alcance

h: Distancia

C= meseta

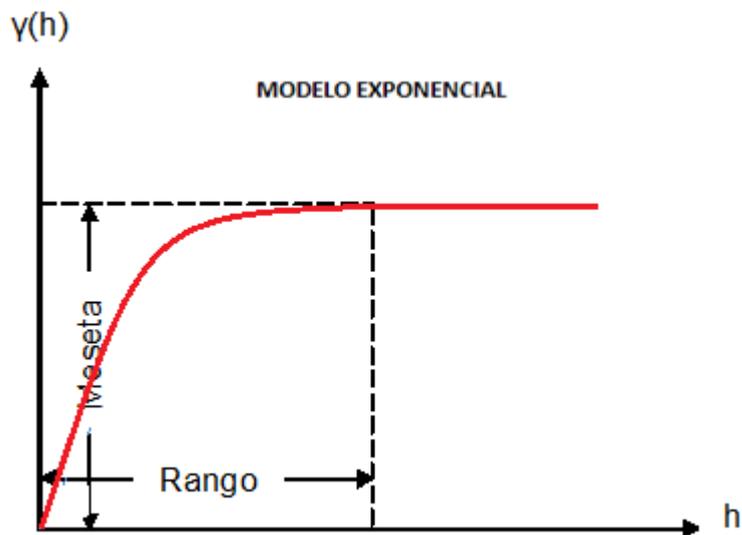


Figura 3.3 Modelo Exponencial
Elaborado por: Dennis Lucin

3.6.4 1.7.2.4 MODELO GAUSSIANO

Este es un modelo continuo, inicialmente presenta un comportamiento parabólico en el origen, después al igual que en el modelo Exponencial alcanza la meseta de forma asintótica. Constituye el único modelo estacionario con un punto de inflexión. Este modelo presenta muy poca variación de las propiedades a medida que la distancia aumenta. Su función está dada por:

$$\gamma(h) = C \left[1 - \exp\left(-3 \frac{h^2}{a^2}\right) \right] \quad (18)$$

Dónde:

$\gamma(h)$ = Modelo gaussiano.

C_o = Covarianza

h = distancia

a = alcance

La figura 3.4 muestra la gráfica del variograma para el modelo gaussiano

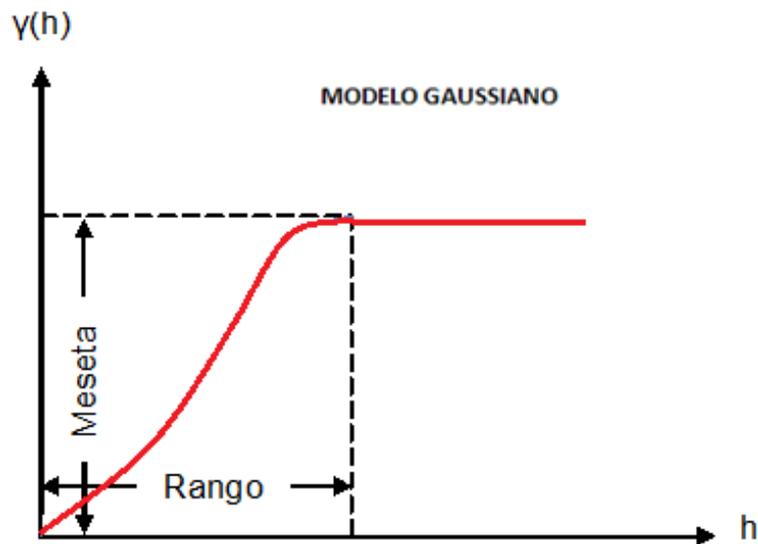


Figura 3.4 Modelo Gaussiano

Elaborado por: Dennis Lucin

3.6.5 MODELO CON EFECTO AGUJERO

El "efectos agujero " es un efecto poco común que se observa cuando el modelado del variograma es en la dirección vertical. Estos variogramas muestran un comportamiento periódico que representa los procesos geológicos cíclicos. Representa modelos continuos cuya función matemática es:

$$\gamma(h) = C \left(1 - \frac{\text{Sen}(ah)}{ah} \right) \quad (19)$$

Dónde:

C: covarianza.

h: distancia.

a: alcance.

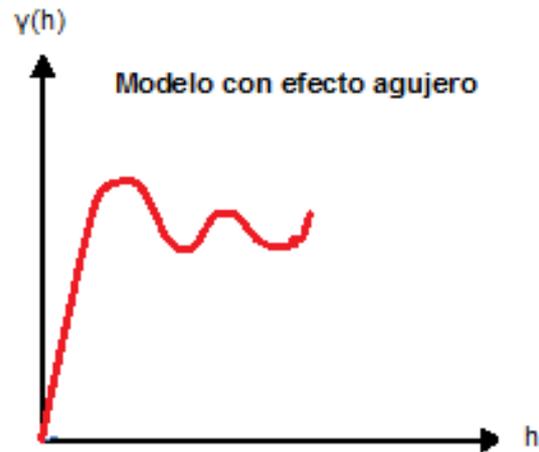


Figura 3.5 Modelo con Efecto Agujero

Elaborado por: Dennis Lucin

3.6.6 MODELO LINEAL

El modelo Lineal no tiene sill ni rango. El sill continúa aumentando mientras la distancia h aumenta, es utilizado para modelar fenómenos que presentan capacidad infinita de dispersión.

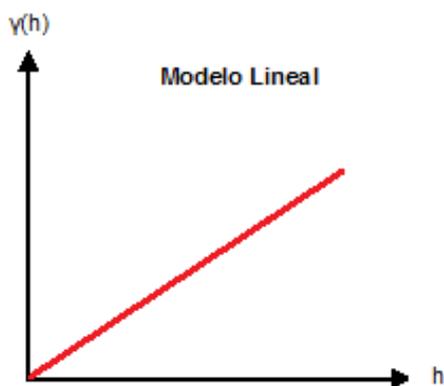


Figura 3.6 Modelo Lineal

Elaborado por: Dennis Lucin

Capítulo IV

4 DESCRIPCIÓN DE VARIOGRAMGENETICS Y DISEÑO DEL ALGORITMO GENÉTICO.

4.1 INTRODUCCION

En el presente capítulo se describirá el desarrollo procedimental y la funcionalidad de VariogramGenetics basada en el ajuste de los variogramas experimental y el modelo teórico.

Además se detallará el modo de operación y ejecución del algoritmo genético que utiliza el VariogramGenetics; este funcionará basado en la interfaz de Microsoft Excel, con todas las funciones disponibles del mismo, así como la edición de gráficos y diferentes objetos que dicho software permita.

En las próximas secciones se describe el diseño específico de cada una de las interfaces que ofrece VariogramGenetics para las diferentes funciones programadas en el mismo, entre ellas, la selección del modelo teórico óptimo para la solución del problema del ajuste variográfico.

La forma de ejecución del algoritmo imita la teoría de la evolución, de tal manera que optimiza el cálculo de variogramas. Para esto se resolvió realizar dichos cálculos en 4 direcciones distintas: 0° , 45° , 90° y 135° . También se tomó en consideración el modelo al cual se ajusta el variograma experimental. Para los modelos teóricos se tomó en cuenta 4 modelos: Modelo Esférico, Modelo Exponencial, Modelo Gaussiano y el Modelo Senoidal.

Para lograr el ajuste variográfico se formuló una función que la denotamos como $V_{func}()$, que es la función objetivo que permite que el algoritmo encuentre el individuo idóneo, resolviendo de esta forma el problema del ajuste variográfico optimizando los cálculos ya mencionados.

Pozos	Coordenada X	Coordenada Y	Valor de la propiedad
Pozo 1	317628,57	9981183,06	0,16812
Pozo 2	316858,54	9985992,9	0,13601
Pozo 3	313800,21	9975558,1	0,16279
Pozo 4	316673,81	9976174,8	0,23652
Pozo 5	316573,88	9974195,44	0,20073
Pozo 6	316445,67	9972197,07	0,1749
Pozo 7	316639,76	9978202,55	0,20668
Pozo 8	317613,09	9979178,37	0,19938
Pozo 9	315571,97	9979157,98	0,19607
Pozo 10	314567,85	9978184,61	0,14539
Pozo 11	315506,76	9977296,69	0,16966
Pozo 12	317693,38	9983193,96	0,17327
Pozo 13	315859,38	9982282,21	0,16824
Pozo 14	315581,18	9981533,61	0,18384
Pozo 15	314582,74	9976192,86	0,19553
Pozo 16	314620,07	9980159,75	0,18864
Pozo 17	316679,94	9980143,61	0,20712
Pozo 18	315543,6	9975159,36	0,22388
Pozo 19	317724,26	9975135,62	0,16674

Tabla 4.2 Ejemplo del formato de ingreso de información
Elaborado por: Dennis Lucin

La interfaz “PRINCIPAL” consta además de cuatro botones principales (figura 4.2) que permite al usuario realizar los cálculos principales y visualización de la distribución de los pozos en el campo; cada uno de estos botones tiene las siguientes funciones:



Figura 4.2 Botones de la Interfaz principal
Elaborado por: Dennis Lucin

El botón “Generar distancias y direcciones”, una vez ingresados los datos, tiene como funcionalidad leer la información, crear las distancias existentes entre los pozos del campo y calcular su dirección, es decir la distancia y el ángulo que forma un pozo con respecto a otro, tomando como referencia el

norte como 0°. De la misma forma genera la variabilidad de la propiedad que presenta los pozos a la cual se está efectuando el cálculo.

Si el algoritmo no es capaz de procesar información, es decir el usuario no ha realizado ingreso de datos, éste emitirá un cuadro de mensaje con el texto “Aún no ha ingresado datos de pozo” (figura 4.3).

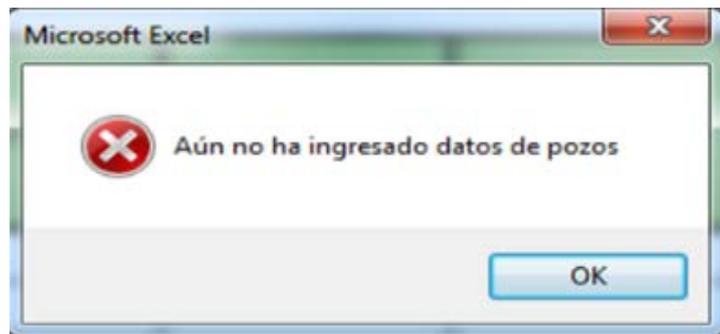


Figura 4.3 Error de ingreso de datos 1
Elaborado por: Dennis Lucin

Si el algoritmo procesa información incompleta, es decir que algún dato de Nombre de pozo, Coordenada en X o Y, o Valor de la propiedad no se ha ingresado, VariogramGenetics emitirá un cuadro de mensaje que con el texto “El número de pozos no coinciden con el número de coordenadas y propiedades” (figura 4.4).

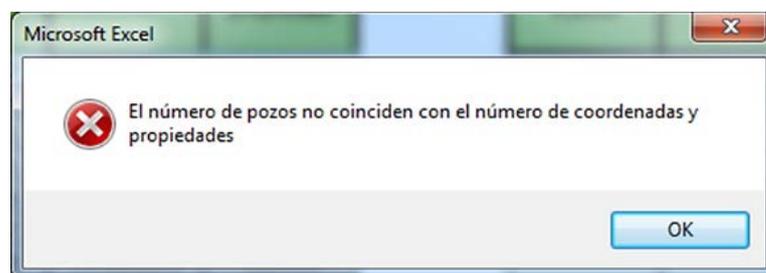


Figura 4.4 Error de ingreso de datos 2
Elaborado por: Dennis Lucin

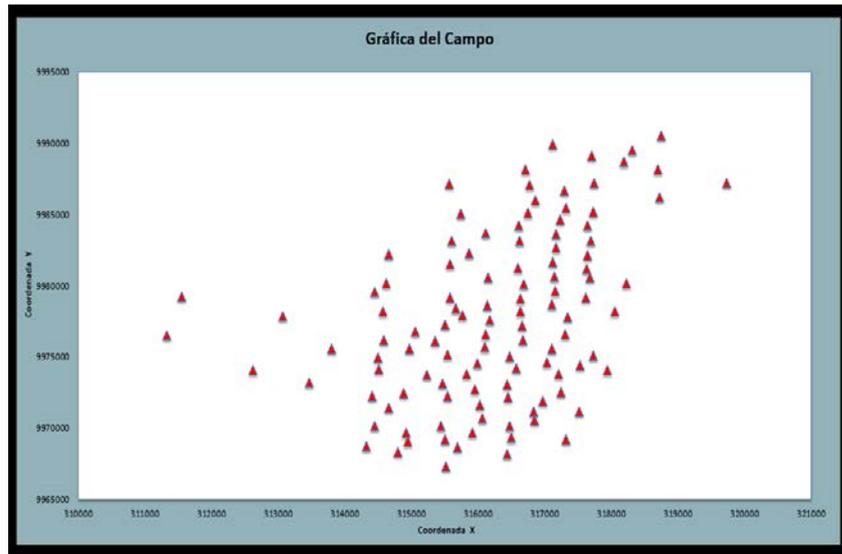


Figura 4.6 Interfaz Ubicación espacial de los pozos
Elaborado por: Dennis Lucin

El botón “Borrar” tiene como única finalidad borrar todos los datos que se hayan calculado o ingresado para el posterior ingreso de nueva información que se desee calcular. Por último tenemos el botón “Ingreso de Parámetros del Algoritmo” nos traslada a una nueva interfaz donde se debe ingresar los distintos parámetros que el algoritmo requiere para su posterior ejecución (Figura 4.7).

Figura 4.7 Interfaz “Ingreso de Parámetros del Algoritmo”
Elaborado por: Dennis Lucin

4.2.1 PARÁMETROS PARA EL CÁLCULO DE VARIOGRAMAS

La interfaz “Ingreso de Parámetros del Algoritmo” mencionada anteriormente, corresponde a la sección donde se realiza la entrada de datos para el algoritmo, valores con la que realizarán los cálculos correspondientes al variograma, como los valores de arranque del algoritmo. En la parte superior de la figura 4.7 se encuentra el bloque perteneciente al ingreso de los límites de los parámetros para el cálculo de variogramas.

Estos parámetros están comprendidos entre un rango máximo y mínimo. Estos límites es la parte significativa e importante de VariogramGenetics, ya que contienen los rangos de valores que el algoritmo optimizará a medida que el software esté en ejecución.

Las 6 variables que se optimizarán están distribuidas de la siguiente manera:

Las primeras cuatro variables (Valor del Lag, Tolerancia Angular, Tolerancia Lag y Ancho de Banda) corresponden a los parámetros que efectuarán el cálculo del Variograma Experimental, mientras que las dos últimas variables (el Sill y Alcance) corresponden a los parámetros que calcularán el Modelo Teórico.

VariogramGenetics proporciona por defecto valores que el usuario podrá tomar a consideración para la ejecución del algoritmo (figura 4.8); dichos valores dependen de la cantidad, distribución y propiedad de cada pozo en el campo. Si el usuario considera estos valores como inadecuados para el caso pertinente, los límites tendrán que ser ingresados manualmente.

Límites de los Parámetros para el cálculo de Variogramas		
	Mínimo	Máximo
Valor del Lag	701,5103829	1044,802697
Tolerancia Angular	15	35
Tolerancia Lag	35	60
Ancho de Banda	746,2876413	7806
Sill	0,001	1
Alcance	746,2876413	11709

Figura 4.8 Ejemplo de valores de límites para el cálculo de variogramas generado por VariogramGenetics
Elaborado por: Dennis Lucin

4.2.2 PARÁMETROS DE ARRANQUE DEL ALGORITMO

Ésta sección se encuentra en la parte inferior de la interfaz “Ingreso de Parámetros del Algoritmo” (figura 4.9). Para el funcionamiento interno del algoritmo, se requerirá en gran medida de estos valores ya que es la encargada de ayudar a reproducir el proceso de evolución de los parámetros anteriores mediante la optimización.

Figura 4.9 Ubicación del bloque Parámetros de Arranque del Algoritmo dentro de la interfaz “Ingreso de Parámetros del Algoritmo”.

Elaborado por: Dennis Lucin

Las variables que se consideran en este bloque son las siguientes:

- **Tamaño de la población:** Indica la cantidad de individuos que el algoritmo creará estructurados con los seis valores de las variables ingresadas en la sección 4.2.1 (Valor del Lag, Tolerancia Angular, Tolerancia Lag y Ancho de Banda, Sill, Alcance) comprendido dentro de su límites correspondientes.
- **Número de Generaciones:** Indica la cantidad de iteraciones máximas que se realizará sobre una población para lograr el proceso evolutivo. El algoritmo creará nueva población hasta encontrar una solución de optimización de las variables de acuerdo a la función objetivo (*Valfunc* ()).
- **Porcentaje de cruzamiento:** Denota el porcentaje de la población sobre la que efectuará la operación de cruzamiento con el mejor individuo.

- Porcentaje de mutación: Expresa la porción de individuos de la población sobre los que se efectuará la operación de Mutación (VariogramGenetics por defecto efectúa la operación de mutación Tipo Un solo punto).
- Tolerancia de la función objetivo: Indica el valor numérico al cual la función objetivo *Valfunc ()* deberá aproximarse.

De la misma forma que para la sección 4.2.1, VariogramGenetics proporciona por defecto ciertos valores (figura 4.10) que el usuario podrá tomar como recomendación de uso para estos parámetros, caso contrario el usuario realizará los cambios respectivos manualmente.

Parámetros de Arranque del Algoritmo	
Tamaño de la Población	50
Número de Generaciones	50
Porcentaje de Cruzamiento	70
Porcentaje de Mutación	10
Tolerancia de la función Objetiva	0,0001

Figura 4.10 Ejemplo de valores de parámetros de arranque del algoritmo generado por VariogramGenetics

Elaborado por: Dennis Lucin

4.3 DISEÑO DEL ALGORITMO GENÉTICO PARA EL AJUSTE VARIOGRÁFICO.

Dentro de la interfaz descrita anteriormente se encuentran dos botones: “Regresar a Principal” y “Comenzar Algoritmo”. El primer botón nos traslada a la interfaz “Principal”, en caso de que el usuario considere necesario cambiar algún dato de ingreso o visualizar la gráfica de ubicación de los pozos del campo.

El segundo botón lee toda la información ingresada y ejecuta el Algoritmo Genético. La estructura como tal del algoritmo que utiliza VariogramGenetics está basada en el esquema mostrado en la figura 4.11.

El esquema con el que fue diseñado el algoritmo genético es el siguiente:

- Generar la población inicial.
- Evaluar la adaptación de todos los individuos en la población.
- Evaluar la población a un objetivo específico.
- Crear una nueva población efectuando operaciones cruce y mutaciones en los individuos en la que ésta acaba de ser medida.
- Reemplazar la antigua población
- Iterar utilizando la nueva población, hasta que la misma cumpla el objetivo estipulado.

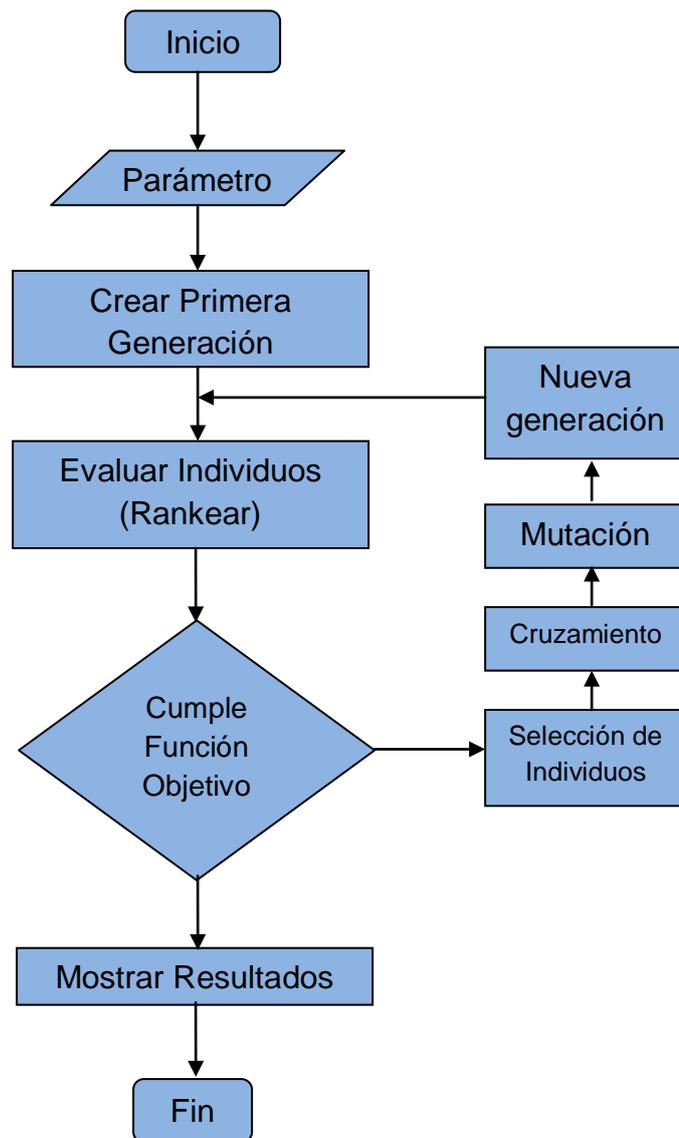


Figura 4.11 Flujograma del Algoritmo Genético

Elaborado por: Dennis Lucin

Luego de la entrada de información el algoritmo este pasa a crear la primera generación (Población Inicial) mediante un sistema aleatorio en forma binaria, es decir caracteres entre 1 y 0. La cantidad de individuos creados depende del valor ingresado en la sección 4.2.1. Cada individuo está dividido en Genes estructurado mediante bits (cantidad de ceros y unos).

VariogramGenetics genera por defecto 6 genes (los parámetros de cálculo de variogramas); cada gen contiene 10 bits (figura 4.12) lo que da un total de 60 bits por individuo (cromosoma) de la población inicial.

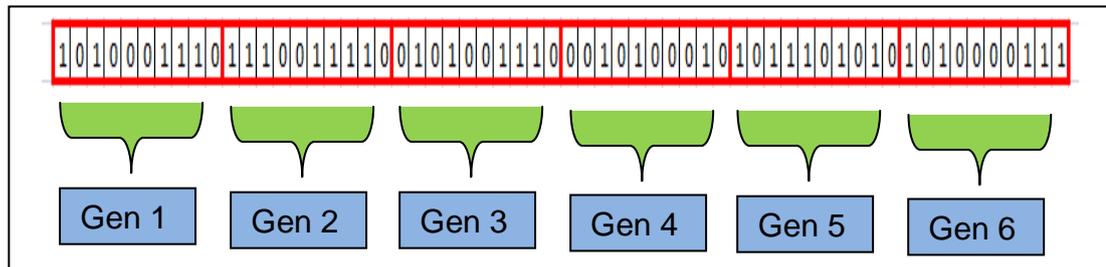


Figura 4.12 Esquema de un individuo de la población inicial

Elaborado por: Dennis Lucin

A continuación se crea una tabla de datos para la lectura de los genes binarios correspondientes a cada cromosoma, es decir, lee el valor binario de cada gen en una tabla generada a partir de los límites ingresados por el usuario en la interfaz de "Ingreso de Parámetros del Algoritmo". La tabla 4.3 muestra un ejemplo de la creación de la tabla de equivalencia.

	Valor del Lag	Tolerancia Angular	Tolerancia Lag	Ancho de Banda	Sill	Alcance
000	100	14	35	50	0,3	100
001	125	20	37	55	0,4	110
010	150	26	39	60	0,5	120
011	175	32	41	65	0,6	130
100	200	38	43	70	0,7	140
101	250	44	45	75	0,8	150
111	300	50	47	80	0,9	160

El diagrama muestra una tabla de equivalencias de código binario con valores reales. La tabla tiene 7 filas y 7 columnas. Las columnas están encabezadas por: Valor del Lag, Tolerancia Angular, Tolerancia Lag, Ancho de Banda, Sill, Alcance. Las filas están encabezadas por los códigos binarios 000, 001, 010, 011, 100, 101, 111. Debajo de la tabla, hay seis cuadros azules etiquetados como Gen 1 a Gen 6, con flechas azules que apuntan hacia arriba a las columnas correspondientes de la tabla. A la derecha de la tabla, hay dos cuadros blancos con flechas rojas que apuntan hacia la izquierda a las columnas de Alcance. El cuadro superior está etiquetado como "Límite Mínimo" y apunta a la fila 000. El cuadro inferior está etiquetado como "Límite Máximo" y apunta a la fila 111.

Tabla 4.3 Ejemplo de Tabla de equivalencias de código binario con valores reales.

Elaborado por: Dennis Lucin

4.3.1 FUNCIÓN OBJETIVO DEL ALGORITMO

El objetivo de VariogramGenetics es tratar de ajustar matemática y gráficamente el variograma experimental con respecto al modelo teórico. De esta forma el algoritmo realizará el ajuste de 4 direcciones distintas para 4 modelos teóricos, es decir que para cada dirección se obtendrá el ajuste para los distintos modelos. Las direcciones a la cual se efectuarán los cálculos son los siguientes:

- Variograma Experimental
 - Variograma en Dirección 0° (N)
 - Variograma en Dirección 45° (N-E)
 - Variograma en Dirección 90° (E)
 - Variograma en Dirección 135° (E-S)

La ecuación que se emplea para realizar el cálculo de variogramas en cualquier dirección, es la ecuación (8) establecida en el capítulo 1, éste permite analizar la correlación de la propiedad en el yacimiento a diferentes distancias.

$$\bar{\gamma}(h) = \frac{\sum (Z(x+h) - Z(x))^2}{2h}$$

Dentro de los distintos modelos teóricos se optó por emplear los siguientes modelos:

- Modelo Teórico
 - Modelo Esférico
 - Modelo Exponencial
 - Modelo Gaussiano
 - Modelo Senoidal

Las ecuaciones de cada modelo a implementarse en el algoritmo genético son las estipuladas en el capítulo 3:

- Modelo Esférico

$$\gamma(h) = C \left[\left(\frac{3h}{2a} \right) - \left(\frac{h^3}{2a^3} \right) \right]$$

- Modelo Exponencial

$$\gamma(h) = C \left[1 - \exp \left(- \frac{3h}{a} \right) \right]$$

- Modelo Gaussiano

$$\gamma(h) = C \left[1 - \exp \left(-3 \frac{h^2}{a^2} \right) \right]$$

- Modelo Senoidal o efecto agujero

$$\gamma(h) = C \left(1 - \frac{\text{Sen}(ah)}{ah} \right)$$

El variograma experimental realiza el cálculo con los parámetros de los primeros cuatro genes (Valor del Lag, Tolerancia Angular, Tolerancia Lag y ancho de banda) una vez que se haya encontrado el valor real en la tabla de equivalencias referente a su valor binario correspondiente al individuo. De éste cálculo se genera el valor de cada lag (distancias) y el valor del variograma a esa distancia ($\gamma(h)_e$) teniendo como resultado el variograma experimental.

Con los valores de cada lag calculados en el variograma experimental, se procede a la obtención de los valores $(\gamma(h)_t)$ correspondiente al modelo teórico; este además trabaja con los valores de los dos últimos genes (Sill, Alcance), utilizando la ecuación que corresponde a cada modelo.

Con los resultados determinados, se procede a evaluar los resultados a una función objetivo. Para efectos de facilidad de comprensión, se decidió establecer la función objetivo, llamada *Valfunc* (), de manera tal que la sumatoria de las diferencias entre los puntos del variograma experimental ($\gamma(h)_e$) y el modelo teórico ($\gamma(h)_t$) sea menor o igual a la tolerancia de la función objetivo (valor ingresado por usuario). De esta forma la ecuación queda establecida así:

$$Valfunc() = \sum (\gamma(h)_e - \gamma(h)_t) \leq Tolerancia$$

La figura 4.13 muestra gráficamente el modo de empleo de la función objetivo. Las flechas (rojas) indican la diferencia que existe entre el variograma experimental (puntos) y el modelo teórico (curva); de esta forma al establecer que la sumatoria de esas diferencias sea menor o igual a un valor de tolerancia ingresada, se ajusten estas dos gráficas resolviendo el problema de ajustamiento variográfico mediante el uso del algoritmo genético. Cabe recalcar que en diversos software esta operación se realiza manualmente de forma no muy precisa.

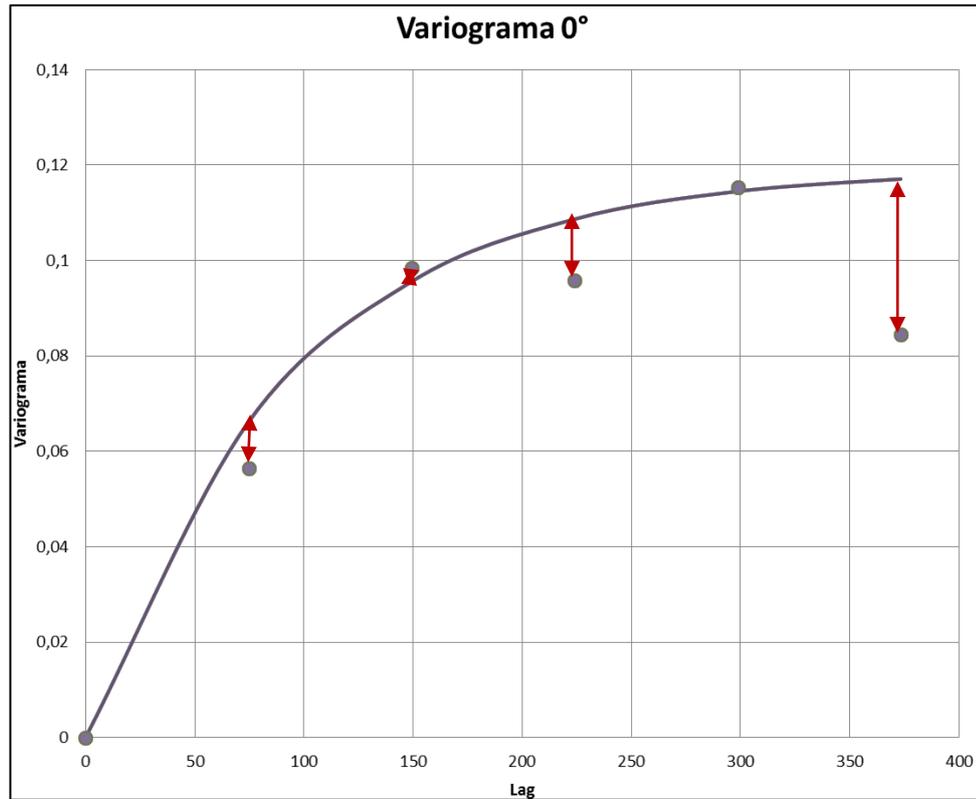


Figura 4.13 Ejemplo de la acción de la función objetiva gráficamente
Elaborado por: Dennis Lucin

Dentro de lo que corresponde el ajuste variográfico, se debe tener en cuenta que los primeros lags (puntos) deben estar mejor ajustado que el restante con la curva del modelo teórico, es decir los lag que se encuentre dentro del Rango o alcance la sumatoria de las diferencias sean mínimas. Debido a esta situación se estableció la multiplicación a la función objetivo por 0.9 a la primera mitad de la sumatoria y la mitad restante multiplicarla por 0.1, de manera que la ecuación determinada para la función objetivo tomaría la siguiente forma:

$$Valfunc() = \left[\sum_{h=1}^{n/2} (\gamma(h)_e - \gamma(h)_t) \right] * 0.9 + \left[\sum_{h=\frac{n}{2}+1}^n (\gamma(h)_e - \gamma(h)_t) \right] * 0.1 \leq Tolerancia$$

Si el valor de la función objetivo (por cada uno de los individuos) no cumple con el requisito establecido, es decir, es mayor a la tolerancia, VariogramGenetics realiza un proceso de Ranqueo, que no es más que un ordenamiento, es decir realiza una categorización en la población desde el mejor individuo hasta el peor dependiendo del valor de su función (de menor a mayor).

4.3.2 OPERACIÓN DE CRUZAMIENTO

El siguiente procedimiento que el algoritmo realiza es la selección aleatoria de un individuo (cromosoma) para efectuar la operación de cruzamiento con el mejor individuo. El cruzamiento es ejecutado sobre un porcentaje de la población (dato de ingreso), que es sometida a cruzarse con el mejor individuo, de esta forma se asegura que la próxima generación pase con mejores características en sus genes y así hallar el individuo que cumpla la función objetivo.

Como se mencionó anteriormente VariogramGenetics realiza solo el cruzamiento del “Tipo un solo punto a nivel de bits”. Para tener una mejor dimensión de cómo se realiza el cruzamiento, supongamos que se han seleccionado los dos individuos que se van a cruzar. Cada uno de estos individuos tiene establecidos sus 6 genes codificados en binario y el valor real (numérico) de cada gen, tomados de la tabla de equivalencia.



La figura 4.14 representa un ejemplo de dos individuos a los que se efectuará el cruzamiento; estos individuos poseen 6 genes de tres bits cada uno (VariogramGenetics trabaja con 6 genes de 10 bits cada uno). Cada color representa un gen de los distintos parámetros y su equivalencia corresponderá de la siguiente manera: Valor del lag (gen 1 amarillo) equivale para este ejemplo a su valor binario multiplicado por 20, Tolerancia angular (gen 2 azul) al valor binario multiplicado por 2, Tolerancia lag (gen 3 verde)

al valor binario multiplicado por 10, Ancho de Banda (gen 4 gris) al valor binario multiplicado por 35, Sill (gen 5 naranja) al valor binario multiplicado por 0.3, Alcance (gen 6 turquesa) al valor binario multiplicado por 25.

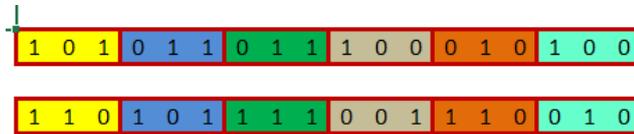


Figura 4.14 Cromosomas (individuos) codificados con parámetros a optimizar.

Elaborado por: Dennis Lucin

De esta forma se obtiene los valores siguientes:

Individuo 1

Valor lag = 100, Tolerancia Angular = 6, Tolerancia lag = 30, Ancho de banda = 140, Sill = 0.6, Alcance = 100

Individuo 2

Valor lag = 120, Tolerancia Angular = 10, Tolerancia lag = 70, Ancho de banda = 35, Sill = 1.8, Alcance = 20

Al efectuarse la operación de cruzamiento de un solo punto a nivel de bits tenemos en la figura 4.15 los hijos resultantes (cruzando desde el bit 8 de los cromosomas del gráfico 4.14, contando siempre de izquierda a derecha).

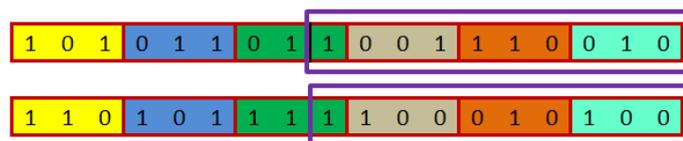


Figura 4.15 Individuos hijos resultantes del cruzamiento de un punto a nivel de bits.

Elaborado por: Dennis Lucin

Hijo 1

Valor lag = 100, Tolerancia Angular = 6, Tolerancia lag = 30, Ancho de banda = 35, Sill = 1.8, Alcance = 20

Hijo 2

Valor lag = 120, Tolerancia Angular = 10, Tolerancia lag = 70, Ancho de banda = 140, Sill = 0.6, Alcance = 100

4.3.3 OPERACIÓN DE MUTACIÓN

En este proceso el programa toma como defecto el tipo de mutación “un solo punto”; esta operación se efectuará a una porción de la población según el valor porcentual que el usuario haya ingresado. Tomando el ejemplo anterior del primer individuo (figura 4.16) tenemos lo siguiente:



Figura 4.16 Cromosoma con parámetros codificados.

Elaborado por: Dennis Lucin

Donde sus valores de los genes son Valor lag = 100, Tolerancia Angular = 6, Tolerancia lag = 30, Ancho de banda = 140, Sill = 0.6, Alcance = 100. El proceso de mutación realiza un cambio en la codificación binaria.

Si la mutación es de un bit (suponiendo que sea el segundo bit del primer gen), se obtendría el cromosoma que muestra el gráfico 4.17.

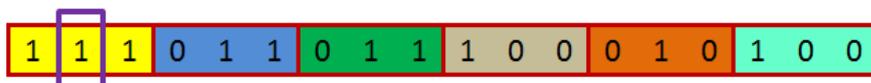


Figura 4.17 Cromosoma afectado por mutación de un bit aleatorio.

Elaborado por: Dennis Lucin

Como se puede apreciar en este caso solo afectó al primer gen dando como resultados Valor lag = 140, Tolerancia Angular = 6, Tolerancia lag = 30, Ancho de banda = 140, Sill = 0.6, Alcance = 100.

4.4 INTERFAZ DE RESULTADOS VARIOGRÁFICOS

Una vez ejecutadas las operaciones de cruzamiento y mutación el algoritmo ingresa la nueva generación a la población y realiza una depuración, es decir elimina los peores individuos. De esta forma vuelve a evaluar la nueva población; si algún individuo cumple la función objetivo, mostrará los resultados obtenidos caso contrario repetirá los procesos de Ranqueo, cruzamiento, mutación y función objetivo hasta conseguir el cromosoma idóneo.

Cabe recalcar que VariogramGenetics realiza el ajuste en cuatro direcciones para cada uno de los modelos (4) mencionados anteriormente. Por lo tanto el algoritmo se ejecutará y determinará 16 respuestas, siendo cada una la representativa (de acuerdo a la tolerancia) para cada método. El software empieza con el variograma 0° y el modelo esférico; una vez encontrado el mejor individuo el algoritmo pasará al siguiente modelo con el mismo variograma y de este modo termina con el último modelo.

Luego de ello comenzará a optimizar los parámetros para el siguiente variograma y modelos hasta llegar al último variograma que es el de 135. Una vez finalizado el ajuste para los cuatro variogramas y modelos el algoritmo pasará a imprimir los resultados de los 25 mejores individuos (por dirección y modelo teórico).

Mientras se ejecuta el algoritmo el usuario podrá visualizar el avance del mismo gráficamente en tiempo real, para esto VariogramGenetics traslada a una nueva interfaz gráfica denominada "Graficas Result.", apenas el algoritmo empiece su ejecución. (Ver figura 4.18).

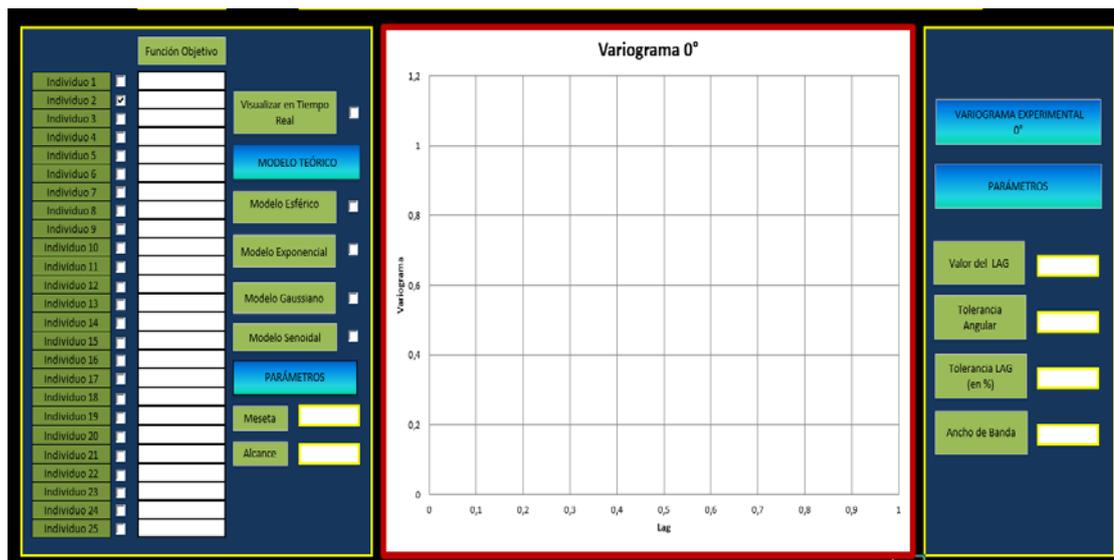


Figura 4.18 Interfaz “Gráficas Result.”

Elaborado por: Dennis Lucin

Esta interfaz está constituida por varias partes, la primera de ellas es la visualización de la información de avance del algoritmo que se encuentra en la parte superior de la interfaz (figura 4.19). En ella se indica en qué variograma (0, 45, 90 o 135), modelo teórico (esférico, exponencial, gaussiano o senoidal) y generación se encuentra el algoritmo, a un determinado instante (actual).

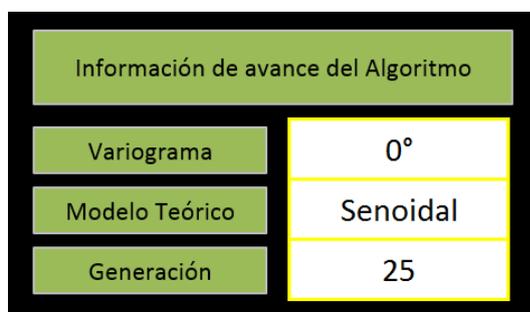


Figura 4.19 Información de Avance del Algoritmo.

Elaborado por: Dennis Lucin

La interfaz además posee una sección de visualización en tiempo real, ésta se activa a través de un "Checkbox", o caja de opción, donde se podrá apreciar el proceso de ajuste que realiza el algoritmo con relación al variograma experimental y el modelo teórico gráficamente (figura 4.20) durante el avance del mismo.

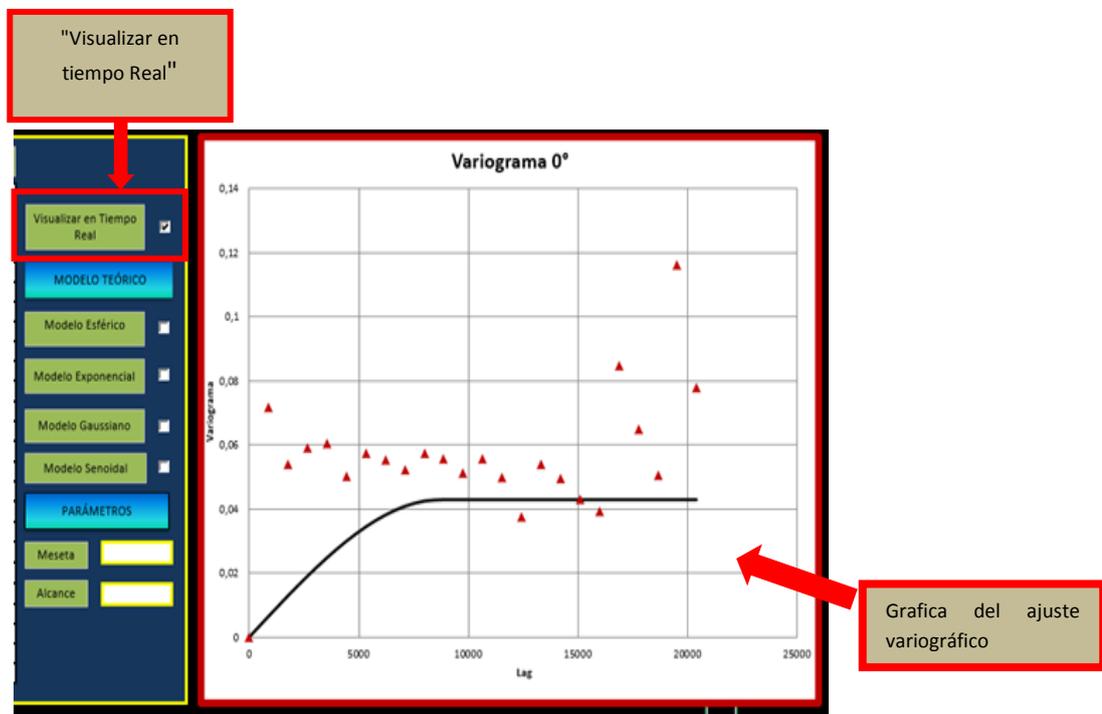


Figura 4.20 Visualización en Tiempo real del ajuste del algoritmo.

Elaborado por: Dennis Lucin

La otra parte de la interfaz es la sección de visualización de resultados de parámetros optimizados (genes) siempre y cuando haya terminado el ajuste variográfico de algún variograma experimental con el método teórico.

Esta sección contiene cada uno de los 25 mejores individuos con sus respectivos valores de la función (valfunc ()), siendo el primero de ellos el individuo idóneo que presenta el mejor ajuste variográfico según la tolerancia ingresada.

La manera de observar gráficamente cada individuo por modelo, es desactivar la opción “Visualizar en tiempo real” para luego hacer click en la caja de opción del individuo que se desee apreciar y el modelo teórico al cual se ajustó el variograma en esa dirección. A continuación se mostrarán cada uno de los parámetros optimizados y su respectiva grafica, estos valores se encuentran en la parte derecha e izquierda de la gráfica, modelo teórico y variograma experimenta respectivamente (grafico 4.21).

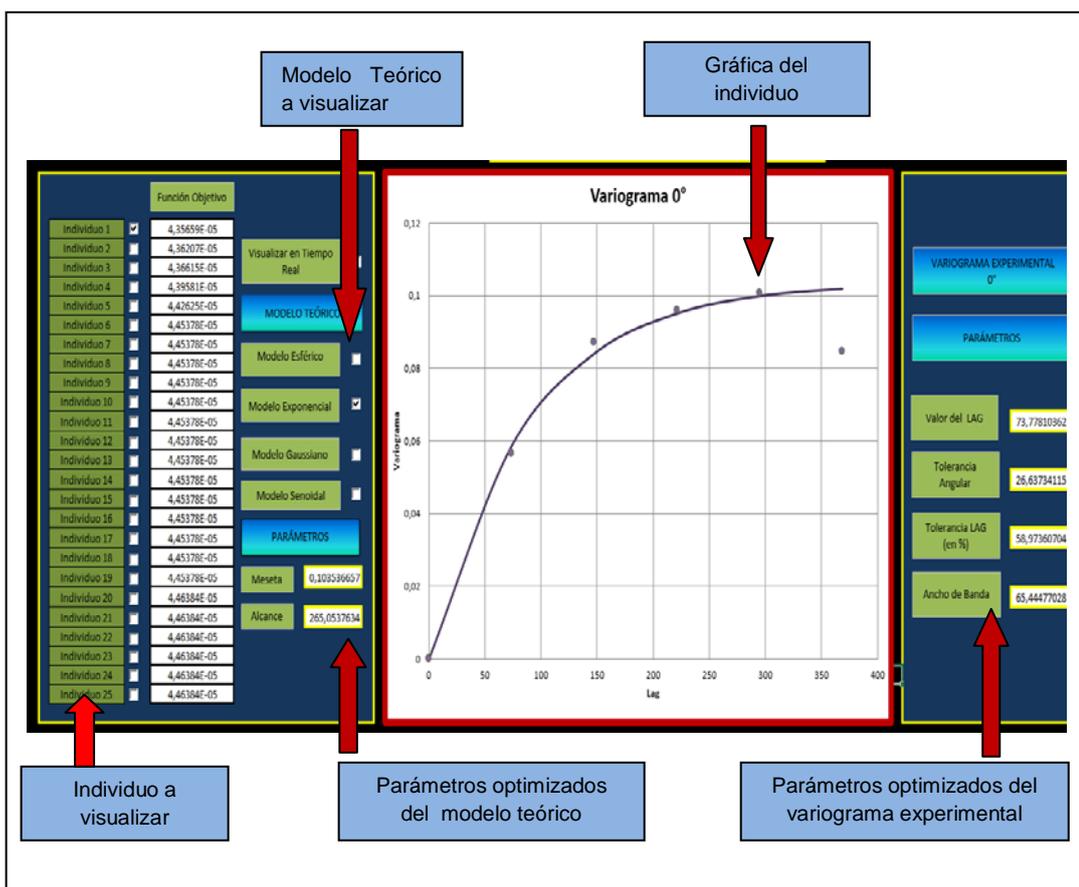


Figura 4.21 Partes de la sección de resultados de la interfaz “Gráficas Result.”

Elaborado por: Dennis Lucin

4.5 RESULTADOS DE LAS ELIPSES DE ANISOTROPÍAS

Una vez que el algoritmo haya terminado, se mostrará una caja de mensaje con el texto “El Algoritmo ha terminado”; de esta forma el usuario podrá tomar decisiones de la elección del mejor ajuste del variograma con su modelo respectivo a una determinada dirección y así poder realizar las elipses de anisotropía que indicarán la tendencia preferencial de la propiedad analizada. VariogramGenetics muestra automáticamente resultados de elipses de anisotropía escogiendo el mejor individuo y graficando para cada dirección el mismo modelo, es decir graficará cuatro elipses por modelo; esta se encuentra en la interfaz “Elipses de anisotropías”.

La figura 4.22 muestra las elipses de anisotropía por modelo teórico con el resultado de las direcciones calculadas (0° , 45° , 90° y 135°), mientras que para las direcciones de (180° , 225° , 270° y 315°) son los mismos valores de las primeras cuatro direcciones, esto se debe a que al ser calculada por la ecuación (8) el resultado será igual ya que son direcciones opuestas.

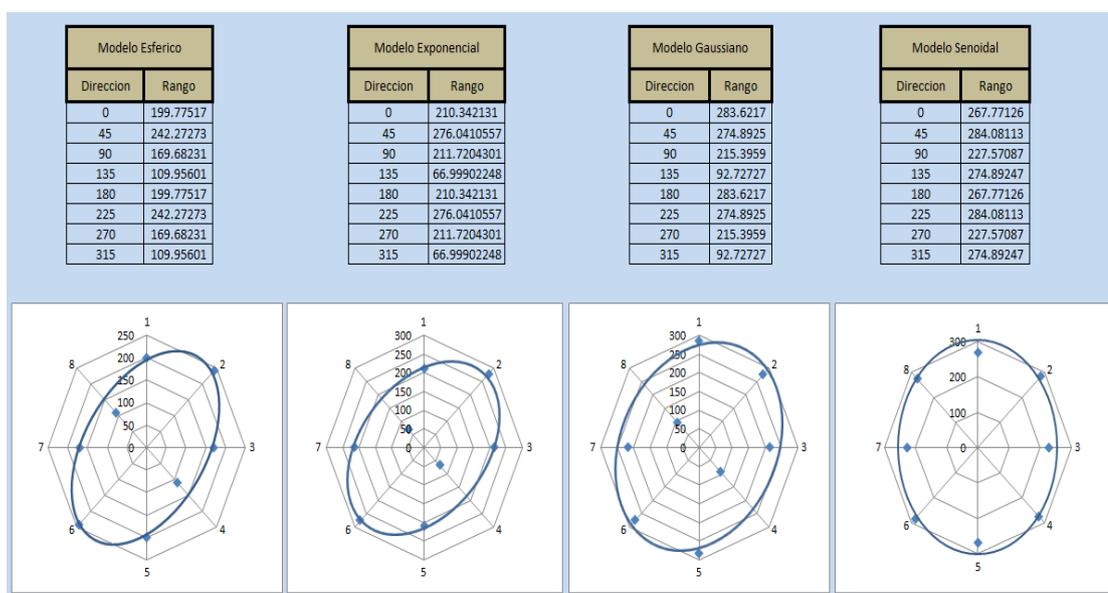


Figura 4.22 Elipses de Anisotropías creados con VariogramGenetics

Elaborado por: Dennis Lucin

Cabe recalcar que el usuario puede tomar a consideración el resultado de las elipses de anisotropía, pero puede tomar resultados de otros ajustes de la interfaz “Grafica Result” según su criterio, y así formar las nuevas elipses de anisotropías.

Capítulo V

5 IMPLEMENTACIÓN DEL DISEÑO DEL ALGORITMO EN EL MODELAMIENTO GEOESTADISTICO

5.1 INTRODUCCIÓN

Para demostrar la funcionalidad de VariogramGenetics se realizó las pruebas para un modelo sintético y para un campo del oriente ecuatoriano que denominaremos campo “UPSE”. El modelo sintético es un modelo ideal generado para realizar pruebas sobre el algoritmo, es decir, un modelo generado con datos conocidos para poder evaluar la efectividad del algoritmo.

Para el caso del modelo sintético se creó pozos que se encuentren distanciados regularmente, es decir pozos equidistantes entre sí, conociendo toda la información necesaria para el algoritmo, incluyendo las respuestas, se podrá determinar si el algoritmo es efectivo o no antes de pasar al caso real.

Para ambos casos de prueba, tanto del modelo sintético como para el campo “UPSE” se eligió analizar como propiedad del yacimiento la porosidad, de esta manera se realizará los cálculos respectivos para optimizar los parámetros o variables y generar la mejor solución, y así graficar la tendencia preferencial de la porosidad en el yacimiento a través de las elipses de anisotropía.

5.2 IMPLEMENTACIÓN DEL ALGORITMO PARA UN CAMPO SINTÉTICO

5.2.1 INFORMACIÓN GENERAL (POZOS, ARENAS)

El modelo sintético contiene un total de 54 pozos. Cada pozo consta del valor correspondiente a la porosidad de una arena específica, para este caso de la “arena A”. La siguiente tabla (5.1) muestra los datos de ingreso que se utilizó en la interfaz “Principal” de VariogramGenetics.

Pozos	Coordenada X	Coordenada Y	Valor de la propiedad
Pozo 1	100	200	20
Pozo 2	100	250	18
Pozo 3	100	300	20
Pozo 4	100	400	21
Pozo 5	100	450	28
Pozo 6	100	500	22
Pozo 7	150	250	16
Pozo 8	150	300	15
Pozo 9	150	400	18
Pozo 10	150	450	25
Pozo 11	150	500	22
Pozo 12	200	200	19
Pozo 13	200	250	14
Pozo 14	200	300	15
Pozo 15	200	350	16
Pozo 16	200	500	19
Pozo 17	250	200	16
Pozo 18	250	250	13
Pozo 19	250	300	19
Pozo 20	250	350	20
Pozo 21	250	400	24
Pozo 22	300	250	14
Pozo 23	300	300	15
Pozo 24	300	350	21
Pozo 25	300	400	21
Pozo 26	300	450	18
Pozo 27	350	200	10
Pozo 28	350	300	19
Pozo 29	350	350	23
Pozo 30	350	400	20
Pozo 31	350	450	18
Pozo 32	350	500	16
Pozo 33	400	150	11
Pozo 34	400	200	12
Pozo 35	400	250	13
Pozo 36	400	300	17
Pozo 37	400	400	13
Pozo 38	400	450	13
Pozo 39	400	500	17
Pozo 40	450	250	10
Pozo 41	450	300	12
Pozo 42	450	350	12
Pozo 43	500	150	18
Pozo 44	500	200	13
Pozo 45	500	250	10
Pozo 46	500	350	11
Pozo 47	500	450	11
Pozo 48	550	150	22
Pozo 49	550	200	19
Pozo 50	550	250	18
Pozo 51	550	300	16
Pozo 52	550	350	17
Pozo 53	550	400	14
Pozo 54	550	450	9

Tabla 5.1 Información de ingreso del modelo sintético.

Elaborado por: Dennis Lucin

Siguiendo con el proceso para poder ejecutar el programa, se procede a calcular todas las distancias, direcciones y variabilidades posibles del campo oprimiendo el botón “Generar distancia y direcciones”. Como resultado se obtuvo un total de 2862 distancias y direcciones posibles en el modelo sintético, donde la mayor distancia presente fue de 570,08 y la menor distancia existente fue de 50. El promedio calculado de las distancias a la que se encuentran los pozos en el campo fue de 50.

Cada uno de los valores obtenidos a consecuencia del primer cálculo se encuentra en la interfaz “Principal” al lado derecho de la tabla de ingreso como se indicó en la sección 4.2. El arreglo de los pozos está distribuido en dirección Este-Oeste como se muestra en la Figura 5.1, cabe resaltar que ésta no necesariamente es la dirección preferencial de la porosidad, es decir la tendencia depositacional (resultado de las elipses de anisotropía) de la arena A.

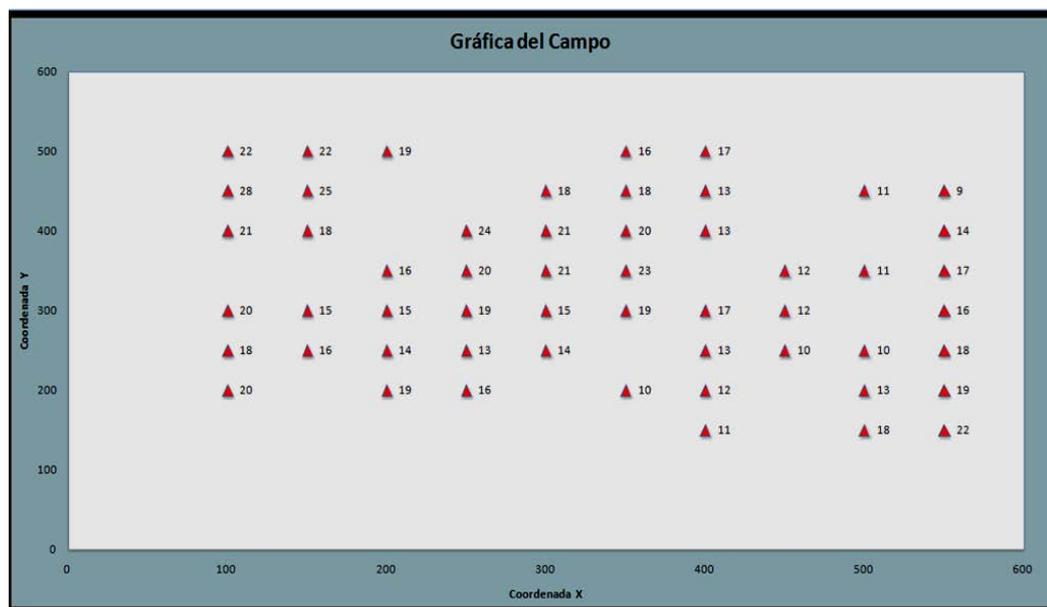


Figura 5.1 Ubicación Espacial de los pozos del modelo sintético.

Elaborado por: Dennis Lucin

A continuación se procede a ingresar los parámetros del algoritmo: los límites para el cálculo de variogramas y para el arranque del algoritmo. Las figuras 5.2 y 5.3 muestran los valores que por defecto VariogramGenetics proporciona para este caso (campo sintético), esto lo realiza automáticamente con el objeto que el usuario parta de datos consistentes.

Límites de los Parámetros para el cálculo de Variogramas		
	Mínimo	Máximo
Valor del Lag	47	70
Tolerancia Angular	15	35
Tolerancia Lag	35	60
Ancho de Banda	50	190
Sill	0,001	1
Alcance	50	285

Figura 5.2 Límites para el cálculo de variogramas proporcionado por VariogramGenetics del modelo sintético.

Elaborado por: Dennis Lucin

Parámetros de Arranque del Algoritmo	
Tamaño de la Población	50
Número de Generaciones	50
Porcentaje de Cruzamiento	70
Porcentaje de Mutación	10
Tolerancia de la función Objetiva	0,00001

Figura 5.3 Parámetros de Arranque proporcionado por VariogramGenetics del modelo sintético.

Elaborado por: Dennis Lucin

Nótese, que el valor promedio de la distancia de los pozos (50), se encuentra dentro de los límites correspondientes al lag con un valor no tan alto para el límite máximo, de esta forma se asegura el programa de tomar la mayor cantidad de puntos para generar el variograma experimental; esto se realiza si se presenta casos donde existan poca cantidad de pozos en el campo, o que en determinada dirección no haya un número suficiente de pozos para correlacionarlos, considerados como una de las problemáticas en el cálculo del variograma experimental.

Los factores más determinantes para que el algoritmo sea eficaz y eficiente son los parámetros de arranque, pero más que todo, la tolerancia de la función objetivo. Para agilizar la velocidad del programa, se estableció que en caso de no haber encontrado una solución, es decir que no se haya ajustado el variograma con la tolerancia establecida y haya corrido todas las generaciones sin respuesta alguna, VariogramGenetics escoja el valor evaluado (del individuo) a la función objetivo de la generación que mejor se ha ajustado, para reemplazarlo con la tolerancia. De ésta forma encontrará el resultado del mejor ajuste para esa dirección.

Por lo tanto se dispuso el cambio de varios parámetros de arranque del algoritmo tales como tamaño de la población a 70, número de generaciones 60 y una tolerancia de 0,000001. Con los parámetros ya ajustados se procedió al arranque del algoritmo presionando el botón “Comenzar Algoritmo”.

Los resultados de cada dirección y por modelo correspondiente al mejor individuo se presentan en la tabla 5.2.

Dirección	Modelo	Valor del Lag	Tolerancia Angular	Tolerancia Lag	Ancho de Banda	Sill	Alcance
Variograma 0°	Esférico	66,73998045	19,32062561	59,97556207	85,71847507	0,108419355	240,8944282
	Exponencial	66,82991202	16,72043011	51,05571848	130,1955034	0,091818182	273,973607
	Gaussiano	64,40175953	29,81915934	39,66764418	74,77028348	0,103536657	181,627566
	Senoidal	69,6627566	16,73998045	57,94721408	134,0273705	0,061545455	68,60703812
Variograma 45°	Esférico	68,9657869	33,02541544	42,40469208	141,0068426	0,064475073	96,63245357
	Exponencial	69,05571848	33,18181818	35,8797654	80,65493646	0,069357771	124,4281525
	Gaussiano	68,67350929	33,31867058	45,72825024	86,12903226	0,066428152	79,17399804
	Senoidal	68,78592375	33,02541544	47,29227761	118,1524927	0,067404692	72,28250244
Variograma 90°	Esférico	53,20527859	20,63049853	39,76539589	189,7262952	0,118184751	285
	Exponencial	63,36754643	26,47605083	59,92668622	115,6891496	0,113302053	284,7702835
	Gaussiano	62,17595308	16,58357771	48,26979472	132,6588465	0,116231672	209,6529814
	Senoidal	64,28934506	29,99511241	59,21798631	130,3323558	0,083029326	152,4535679
Variograma 135°	Esférico	67,18963832	32,90811339	58,72922776	74,63343109	0,134785924	250,7722385
	Exponencial	67,18963832	34,23753666	58,55816227	160,1661779	0,134785924	285
	Gaussiano	67,09970674	35	58,97360704	179,4623656	0,129903226	176,5738025
	Senoidal	67,30205279	33,90518084	59,53567937	102,4144673	0,114278592	139,5894428

Tabla 5.2 Resultados de las optimizaciones del Algoritmo.
Elaborado por: Dennis Lucin

Debido a que la tolerancia resultó ser muy baja (0,000001) el programa eligió el mejor individuo de todas las 70 generaciones por dirección y modelo. De esta forma se obtuvo que el variograma en la dirección 0° , el modelo al que mejor se ajustó fue el esférico con valor de la función objetivo de 0,0000202 (figura 5.4).

Para la dirección 45° el modelo que mejor se ajusta es el modelo exponencial con un valor de la función de 0,000414851 (Figura 5.5). En la dirección 90° el modelo que más coincide con el variograma experimental fue el esférico (figura 5.6) con un valor de la función de 0,0001023 y finalmente el modelo esférico fue el que mejor se ajustó al variograma en la dirección 135° con una diferencia de 0,001375467 (figura 5.7).

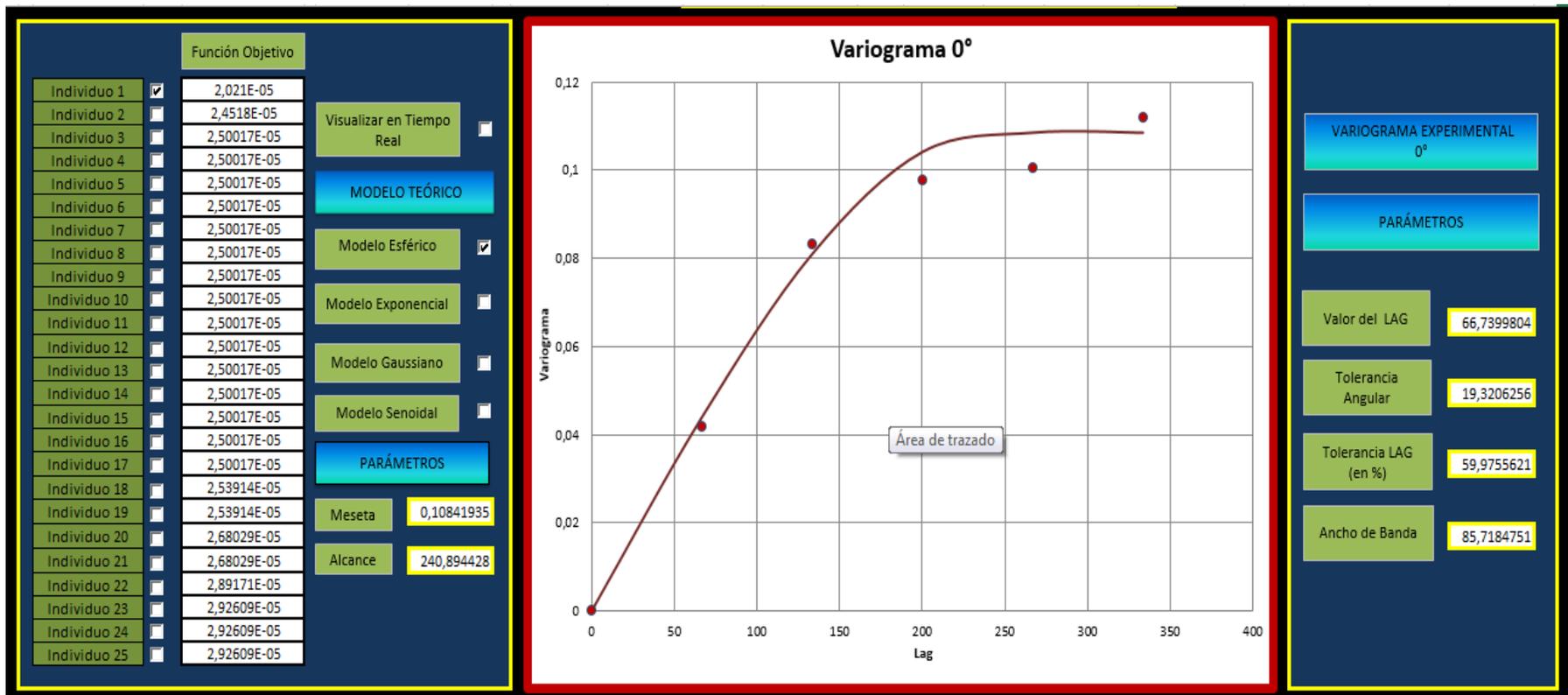


Figura 5.4 Mejor ajuste (Esférico) en el variograma 0°

Elaborado por: Dennis Lucin

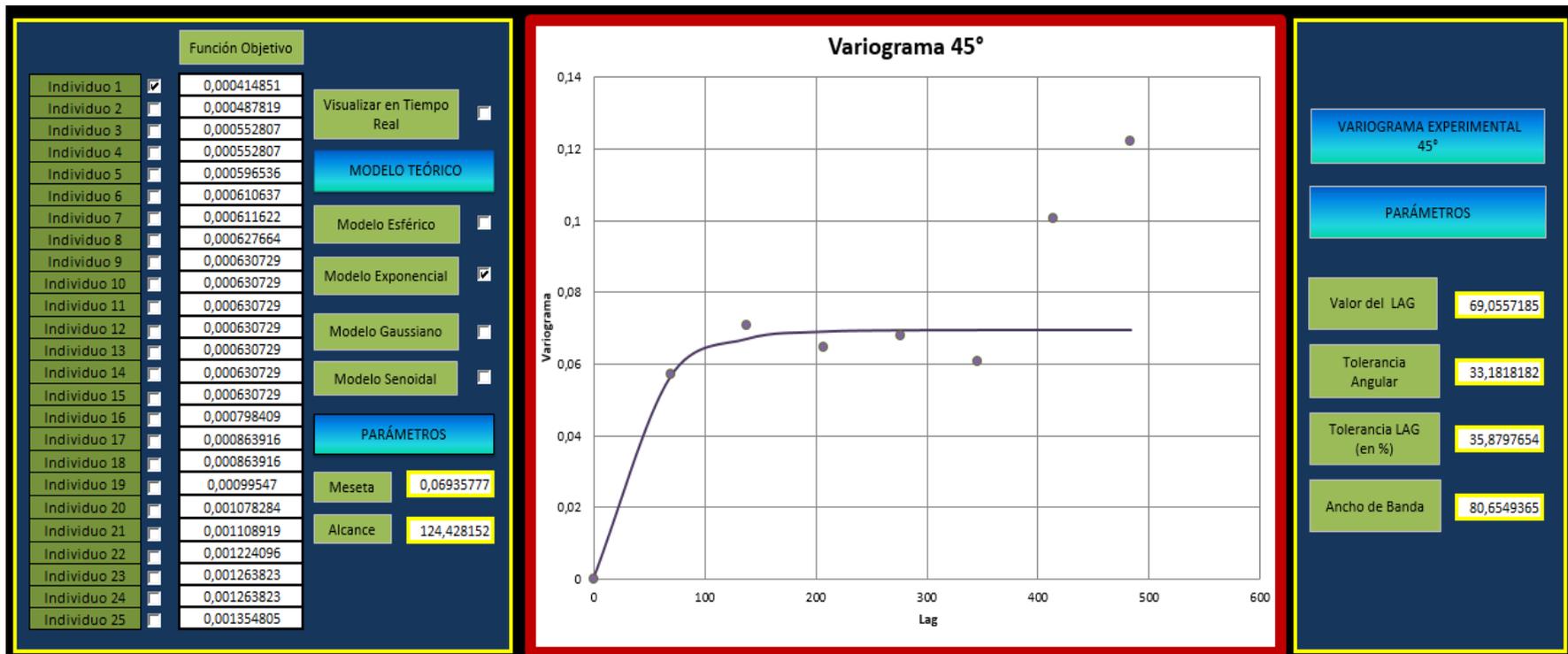


Figura 5.5 Mejor ajuste (Exponencial) en el variograma 45°

Elaborado por: Dennis Lucin

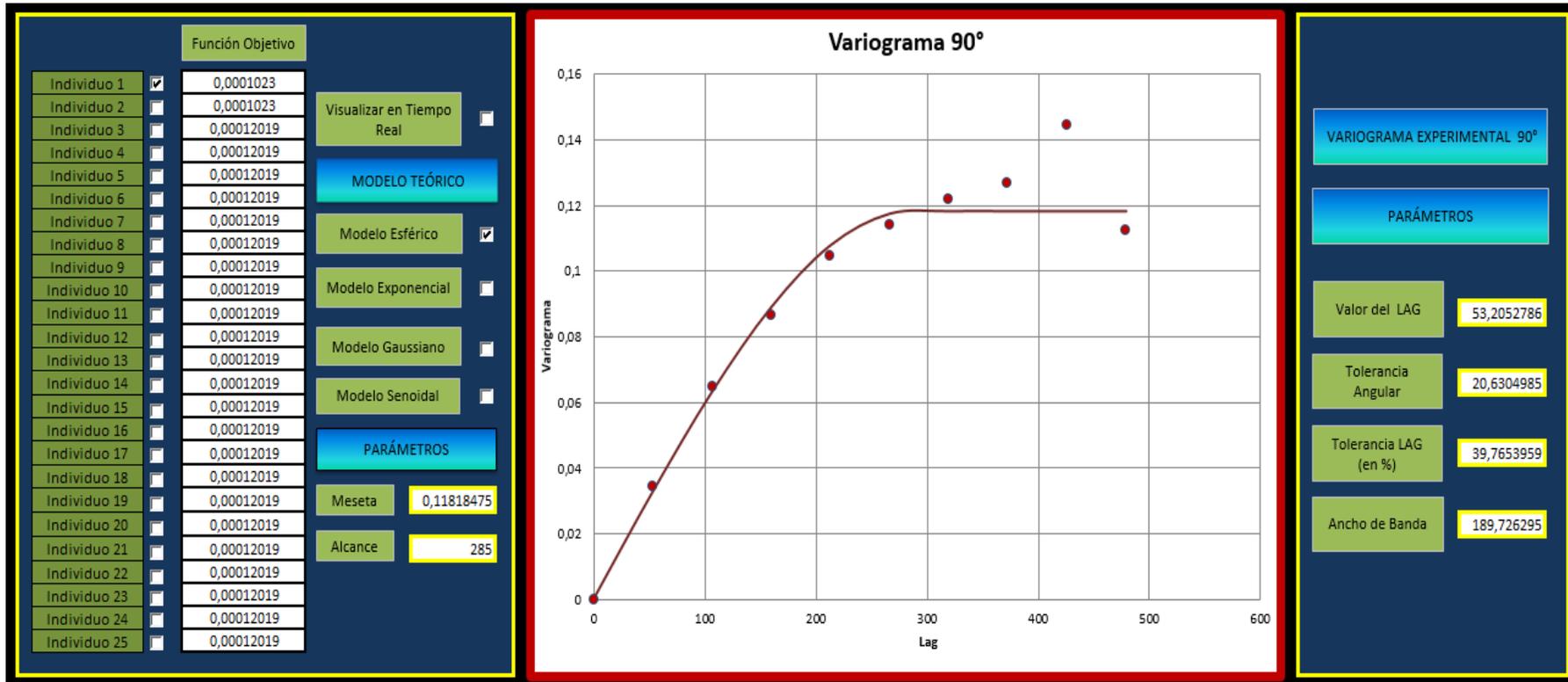


Figura 5.6 Mejor ajuste (Esférico) en el variograma 90°

Elaborado por: Dennis Lucin

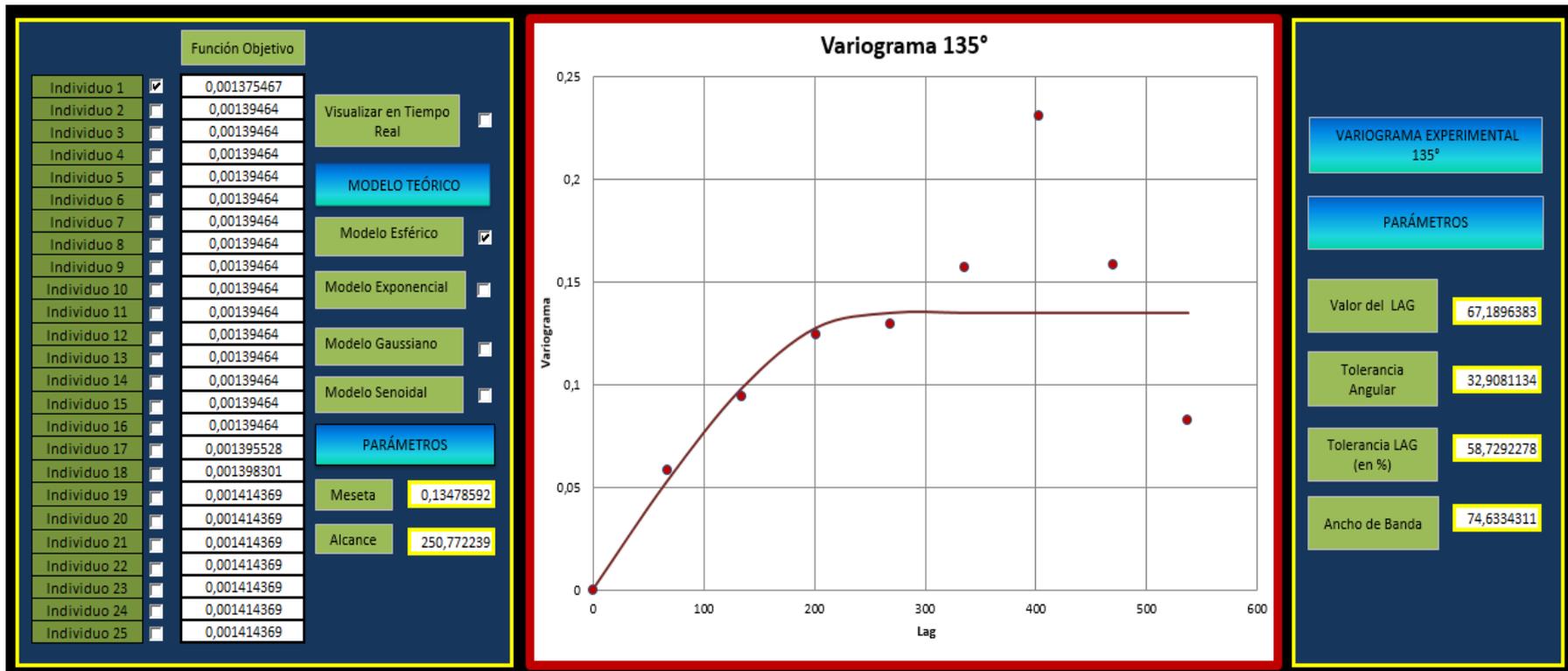


Figura 5.7 Mejor ajuste (Esférico) en el variograma 135°

Elaborado por: Dennis Lucin

5.3 RESULTADOS DE LAS ELIPSES DE ANISOTROPÍA DEL CAMPO SINTÉTICO.

A través del ajuste de los variogramas se obtienen los resultados del alcance y meseta, que son indispensables al momento de construir las elipses de anisotropía; a partir de estos obtenemos parámetros como rango mayor, menor y dirección preferencial de la elipse. Las figuras 5.8 y 5.9 muestran las elipses de anisotropía que VariogramGenetics genera automáticamente tal y como se lo mencionó en el capítulo 4 sección 4.5, a partir del alcance o rango resultante proporcionado por el programa.

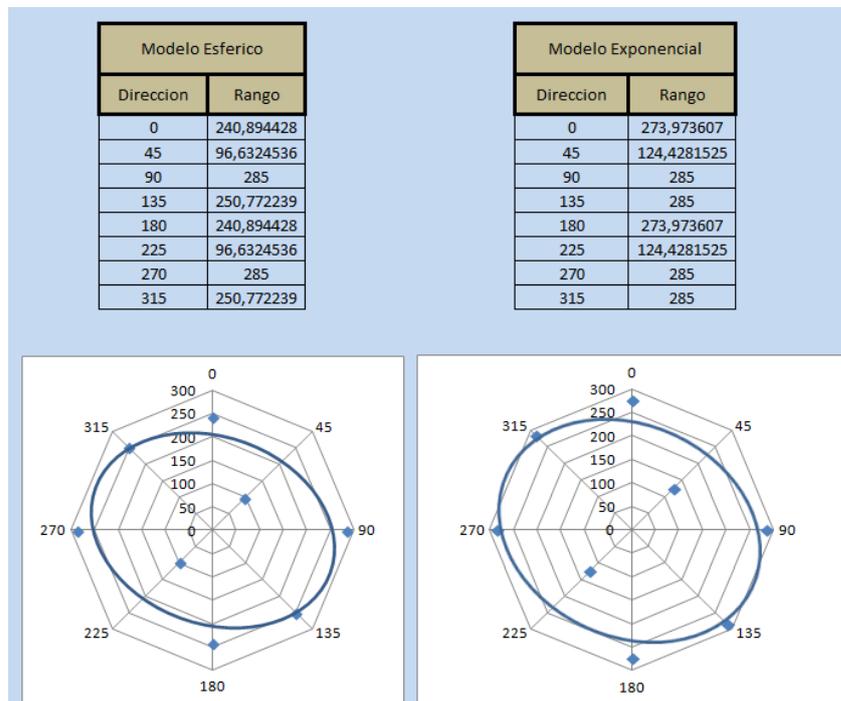
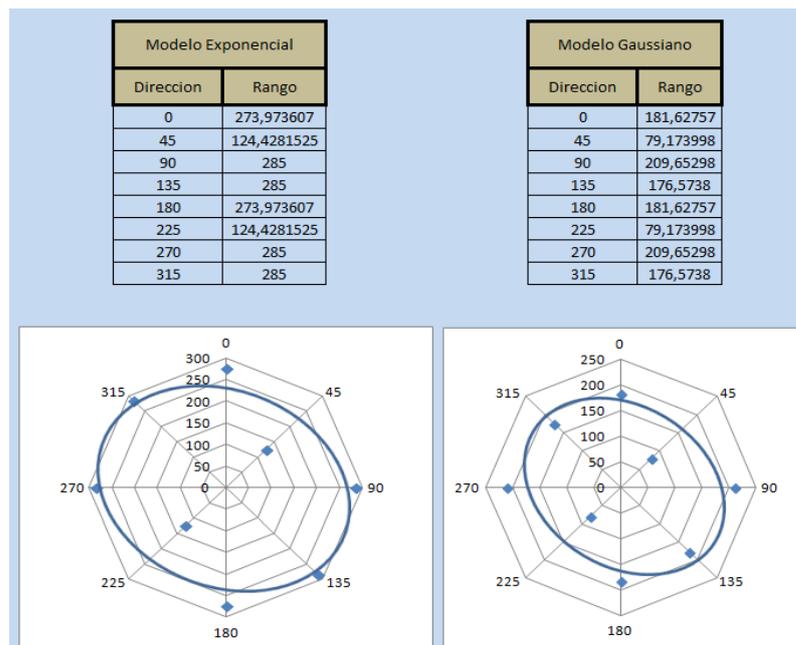


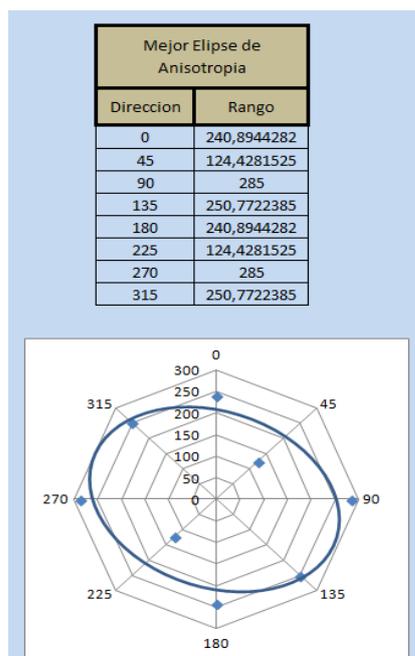
Figura 5.8 Elipses de anisotropía del modelo Esférico y Exponencial del campo sintético.

Elaborado por: Dennis Lucin



**Figura 5.9 Elipses de anisotropía del modelo Gaussiano y Senoidal del campo sintético.
Elaborado por: Dennis Lucin**

Finalmente se obtuvo la elipse de anisotropía con los modelos que mejor se ajustaron a la dirección de 0°, 45°, 90°, y 135° (figura 5.10)



**Figura 5.10 Elipses de anisotropía del mejor ajuste del campo sintético.
Elaborado por: Dennis Lucin**

Según los resultados obtenidos de las elipses de anisotropía, se puede distinguir que la tendencia preferencial de la arena A del campo sintético tiene una disposición en dirección Este-Oeste, es decir que los valores de la propiedad se inclinan en esa dirección (son parecidos). Esto se debe al tipo de ambiente depositacional al cual se constituyó la formación. Debido a que éste es un campo sintético, no se tiene a disposición de un modelo de ambiente depositario y mapas de facies para corroborar los resultados, ya que este análisis se necesita de estudios de núcleos (muestras), pero mediante la visualización de la distribución de la propiedad (porosidad) en el campo se puede apreciar la dirección donde los valores son semejantes entre sí (Ver figura 5.11), de esta forma se observa que la dirección del resultado obtenido por el algoritmo es igual a la dirección de la figura 5.11.

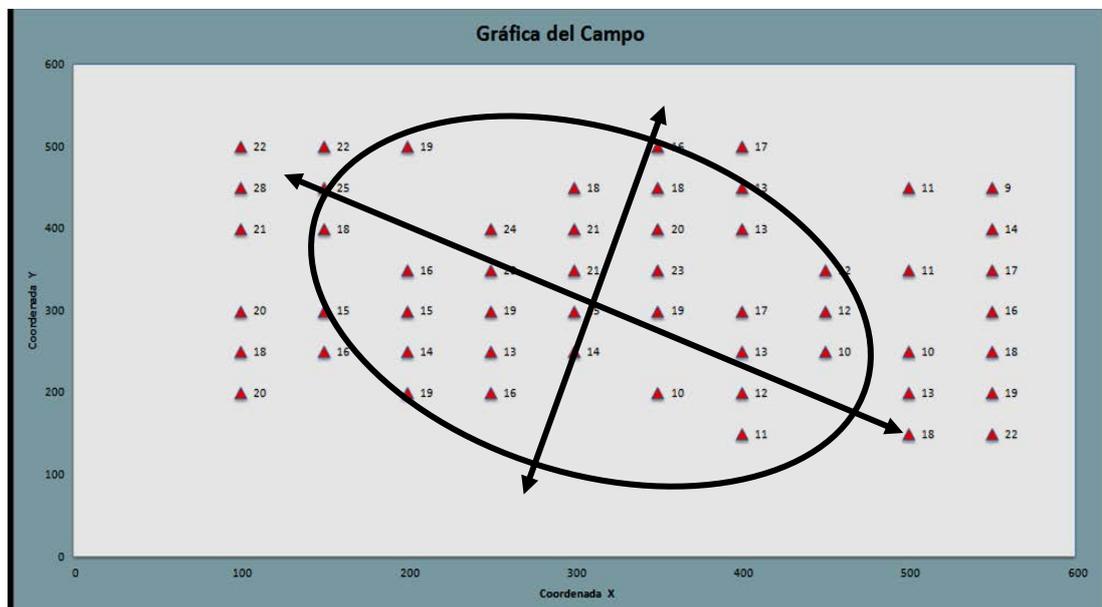


Figura 5.11 Dirección preferencial de la porosidad en modelo sintético.

Elaborado por: Dennis Lucin

5.4 IMPLEMENTACIÓN DEL ALGORITMO AL CAMPO DEL ORIENTE ECUATORIANO.

5.4.1 CAMPO DEL ORIENTE ECUATORIANO

Para demostrar la funcionalidad de VariogramGenetics, se ha escogido un campo perteneciente al oriente Ecuatoriano, al mismo que denominamos anteriormente como campo “UPSE”, del cual se requerirá de toda la información necesaria para el algoritmo y su posterior correlación de la propiedad.

5.4.2 LOCALIZACIÓN

El campo “UPSE” se localiza en la zona nororiente del Ecuador, región de donde se obtiene la mayor producción del país.



Figura 5.12 Región de Localización del campo “UPSE”

5.4.3 TIPO DE AMBIENTE SEDIMENTARIO

Los estudios del análisis sedimentológico realizado al campo UPSE, determinaron que el ambiente de depositación, en general, consiste de un ambiente tipo llanura costera, en la que existen una serie de numerosos canales distributarios, y regiones pantanosas en las que se habría producido la acumulación de carbones. Los sedimentos deformados podrían corresponder a deslizamientos de los flancos de los canales, un rasgo bastante común en este tipo de ambiente sedimentario. Existen también pequeños niveles de areniscas que White et al. (1995 B) interpretan como depósitos de desbordamiento. La presencia de facies de tipo heterolítico hacia techo así como las numerosas truncaciones que se observan en las areniscas corrobora la interpretación de un ambiente costero dominado por marea.

Un análisis realizado al Miembro Arenisca U, efectuado a varios núcleos revelan que éste empieza con estructuras de carga formadas en lodolitas que han sido sobreyacidas abruptamente por arenisca, las mismas que, en la mayor parte, son de tamaño fino a medio con estratificación cruzada variable, con interlaciones de láminas de carbón y "clay drapes". White et al. (1995-B), interpretan estas areniscas como depósitos fluviales de baja sinuosidad, varios autores lo interpretan como depósitos fluviales de tipo meandriforme con influencia de marea, donde pueden estar amalgamadas, también, barras de marea dentro del curso fluvial.

Este depósito demuestra, por correlación entre pozos, para la base del miembro Arenisca U, una geometría elongada en dirección Este-Oeste y una distribución areal limitada, por los que se los ha interpretado como barras de arena y canales mareales que se desarrollaron dentro de un estuarino protegido del mar abierto por una isla barrera. Las figuras 5.13 y 5.14 presentan un análogo moderno de un estuarino para la arena U del campo UPSE.



Figura 5.13 Sistema deposicional análogo al Miembro Arenisca U para el campo UPSE.

Fuente: Diana Ramírez, 2007

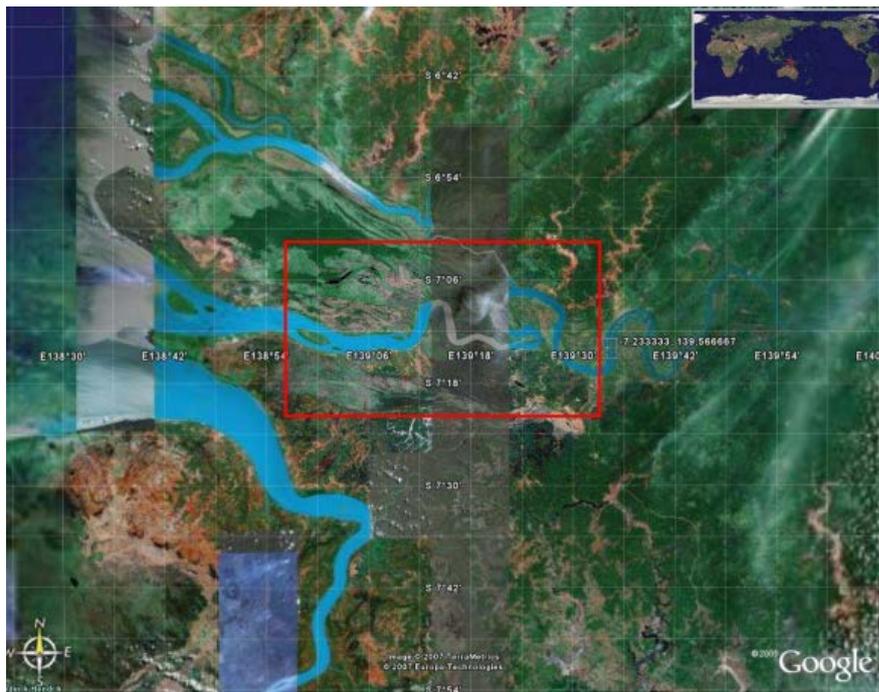


Figura 5.14 Sistema deposicional dominado por corrientes fluviales análogo al Miembro Arenisca U para el campo UPSE.

Fuente: Diana Ramírez, 2007

5.4.4 INFORMACIÓN GENERAL

El campo “UPSE” tiene un total de 106 pozos, produce 45153 barriles diarios en 97 pozos productores; la producción viene principalmente de las arenas U inferior y T inferior, pero se pueden encontrar pozos que producen de las arenas Basal Tena, U superior y T Superior, a pesar que son muy pocos. La arena en la que se le realizará el análisis variográfico usando el software es la arena U inferior.

En la siguiente tabla (5.3) se muestran algunos de los datos de ingreso en la interfaz “Principal” de VariogramGenetics del campo propuesto.

Pozos	Coordenada X	Coordenada Y	Valor de la propiedad
Pozo 1	317628,57	9981183,06	0,16812
Pozo 2	316858,54	9985992,9	0,13601
Pozo 3	316673,81	9976174,8	0,23652
Pozo 4	316573,88	9974195,44	0,20073
Pozo 5	316445,67	9972197,07	0,1749
Pozo 6	316639,76	9978202,55	0,20668
Pozo 7	317613,09	9979178,37	0,19938
Pozo 8	315571,97	9979157,98	0,19607
Pozo 9	314567,85	9978184,61	0,14539
Pozo 10	315506,76	9977296,69	0,16966
Pozo 11	317693,38	9983193,96	0,17327
Pozo 12	315859,38	9982282,21	0,16824
Pozo 13	315581,18	9981533,61	0,18384
Pozo 14	314582,74	9976192,86	0,19553
Pozo 15	314620,07	9980159,75	0,18864
Pozo 16	316679,94	9980143,61	0,20712
Pozo 17	315543,6	9975159,36	0,22388
Pozo 18	317724,26	9975135,62	0,16674
Pozo 19	314504,61	9974181,42	0,20419
Pozo 20	315466,71	9973162,27	0,1815
Pozo 21	315502,49	9969177,53	0,17895
Pozo 22	317518,32	9971184,84	0,17611
Pozo 23	316473,89	9970129,55	0,16562
Pozo 24	314447,42	9970191,09	0,18219
Pozo 25	314408,71	9972234,41	0,15068
Pozo 26	315607,99	9983145,78	0,14598

Tabla 5.3 Información de ingreso del campo “UPSE”

Elaborado por: Dennis Lucin

Se calculó a continuación todas las distancias, direcciones y variabilidades posibles del campo oprimiendo el botón “Generar distancia y direcciones”. Se obtuvo un total de 11130 distancias y direcciones posibles en el campo, donde la mayor distancia presente fue de 23418 metros y la menor distancia existente fue de 483 metros. El promedio calculado de las distancias a la que se encuentran los pozos en el campo es de 696,35 metros.

Cada uno de los valores conseguidos derivados del primer cálculo se localizan en la tabla contigua a la tabla de ingreso de la interfaz “Principal” tal y como se indicó en la sección 4.2. El arreglo distributivo de los pozos se encuentra en una dirección NE-SO (Figura 5.15).

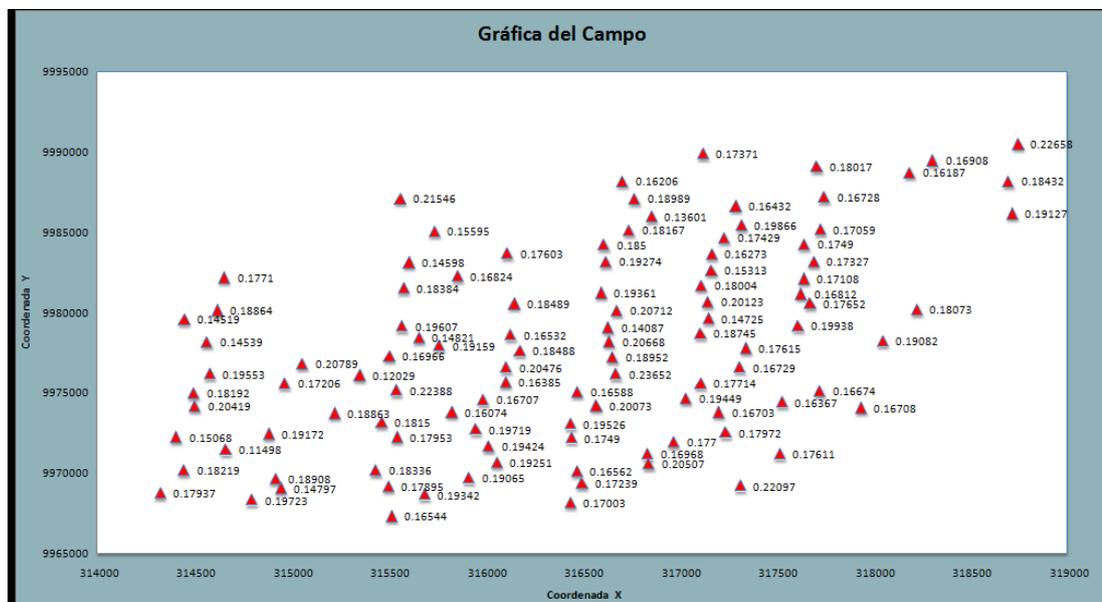


Figura 5.15 Ubicación Espacial de los pozos del Campo “UPSE”.

Elaborado por: Dennis Lucin

En la fase de ingreso de los parámetros del algoritmo, tanto los límites para el cálculo de variogramas como para el arranque del algoritmo, se muestran en las figuras 5.16 y 5.17 que por defecto VariogramGenetics proporciona para este campo.

Límites de los Parámetros para el cálculo de Variogramas		
	Mínimo	Máximo
Valor del Lag	654.5722461	974.8948347
Tolerancia Angular	15	35
Tolerancia Lag	35	60
Ancho de Banda	.353453387411	7806
Sill	0,001	1
Alcance	696.3534533	11709

Parámetros de Arranque del Algoritmo	
--------------------------------------	--

Figura 5.16 Límites para el cálculo de variogramas proporcionado por VariogramGenetics para el campo “UPSE”.

Elaborado por: Dennis Lucin

Sill	0,001	1
Alcance	696.3534533	11709

Parámetros de Arranque del Algoritmo	
Tamaño de la Población	50
Número de Generaciones	50
Porcentaje de Cruzamiento	70
Porcentaje de Mutación	10
Tolerancia de la función Objetiva	0,00001

Figura 5.17 Parámetros de Arranque proporcionado por VariogramGenetics para el campo “UPSE”.

Elaborado por: Dennis Lucin

De igual forma que en el modelo sintético, para el campo “UPSE” se puede notar que el valor promedio de la distancia de los pozos (696,35 metros) se localiza entre los límites del valor del lag (ver figura 5.16), valores que el software proporciona como datos consistentes para el usuario.

Se decidió trabajar con límites del valor de lag de 654,57 a 800. Para el arranque del algoritmo se modificó el número de generaciones a 60, porcentaje de cruzamiento a 65, porcentaje de mutación a 8 y una tolerancia de 0,0001. Con los parámetros ya ingresados se procedió al arranque del algoritmo presionando el botón “Comenzar Algoritmo”.

Una de las formas de agilizar la velocidad del programa, fue la implantación al algoritmo en caso de no haber encontrado una solución, es decir que no se haya ajustado el variograma con la tolerancia establecida y una vez corridas todas las generaciones, escoja el valor de la función (del individuo) en la generación que mejor se ha ajustado y tomarla como la mejor respuesta.

Los resultados de cada dirección y por modelo del mejor individuo se presentan en la tabla 5.4.

Dirección	Modelo	Valor del Lag	Tolerancia Angular	Tolerancia Lag	Ancho de Banda	Sill	Alcance
Variograma 0°	Esférico	795,4509403	33,90518084	59,90224829	3150,231785	0,056662757	741,019384
	Exponencial	796,4460471	35	59,07135875	3094,306641	0,057639296	654,5722462
	Gaussiano	795,7352565	33,57282502	59,87781036	3136,250499	0,054709677	654,5722462
	Senoidal	773,8429064	33,514174	56,5542522	3136,250499	0,057639296	730,2134918
Variograma 45°	Esférico	763,7496801	17,79569892	45,26392962	5408,209463	0,054709677	903,1077675
	Exponencial	759,6270947	17,81524927	44,67741935	6680,506483	0,057639296	741,019384
	Gaussiano	766,1663681	17,6001955	48,46529814	6687,497126	0,056662757	773,4370607
	Senoidal	761,7594665	17,26783969	45,04398827	2220,47627	0,055686217	7937,743611
Variograma 90°	Esférico	757,2104067	16,58357771	57,53176931	2514,083275	0,066428152	1043,584367
	Exponencial	738,8720096	31,20723363	36,36852395	3856,286724	0,068381232	1011,16669
	Gaussiano	793,3185685	23,64125122	48,80742913	6652,543911	0,063498534	1508,237733
	Senoidal	761,0486759	32,41935484	58,02052786	4010,08087	0,055686217	11222,73485
Variograma 135°	Esférico	795,166624	20,21994135	46,0459433	731,4693188	0,063498534	1259,702211
	Exponencial	792,749936	20,76735093	47,07233627	7393,552065	0,060568915	1119,225612
	Gaussiano	795,0244659	20,88465298	45,87487781	4604,285522	0,063498534	1021,972582
	Senoidal	740,0092746	20,72825024	49,34506354	2220,47627	0,057639296	892,3018753

**Tabla 5.4 Resultados de las optimizaciones del Algoritmo del campo “UPSE”.
Elaborado por: Dennis Lucin**

Debido a que la tolerancia resultó ser muy baja (0,0001) el programa no pudo obtener una solución óptima para el variograma de 0° y 135° bajo la tolerancia ingresada, así que eligió la respuesta al que mejor se ajustó en esa dirección. De esta forma se obtuvo que el variograma en la dirección 0°, el modelo que mejor se ajustó fue el Esférico con valor de la función objetivo de 0,000240527 (figura 5.18).

Para la dirección 45° el modelo que mejor se ajusta es el modelo Exponencial con una diferencia entre las dos gráficas de 9,68206E-05 (Figura 5.19). En la dirección 90° el modelo que más coincide con el variograma experimental fue el Esférico (figura 5.20) con un valor de la función de 4,00711E-05 y finalmente el modelo Exponencial fue el que mejor se ajustó al variograma en la dirección 135° con una diferencia de 0,000145761 (figura 5.21).

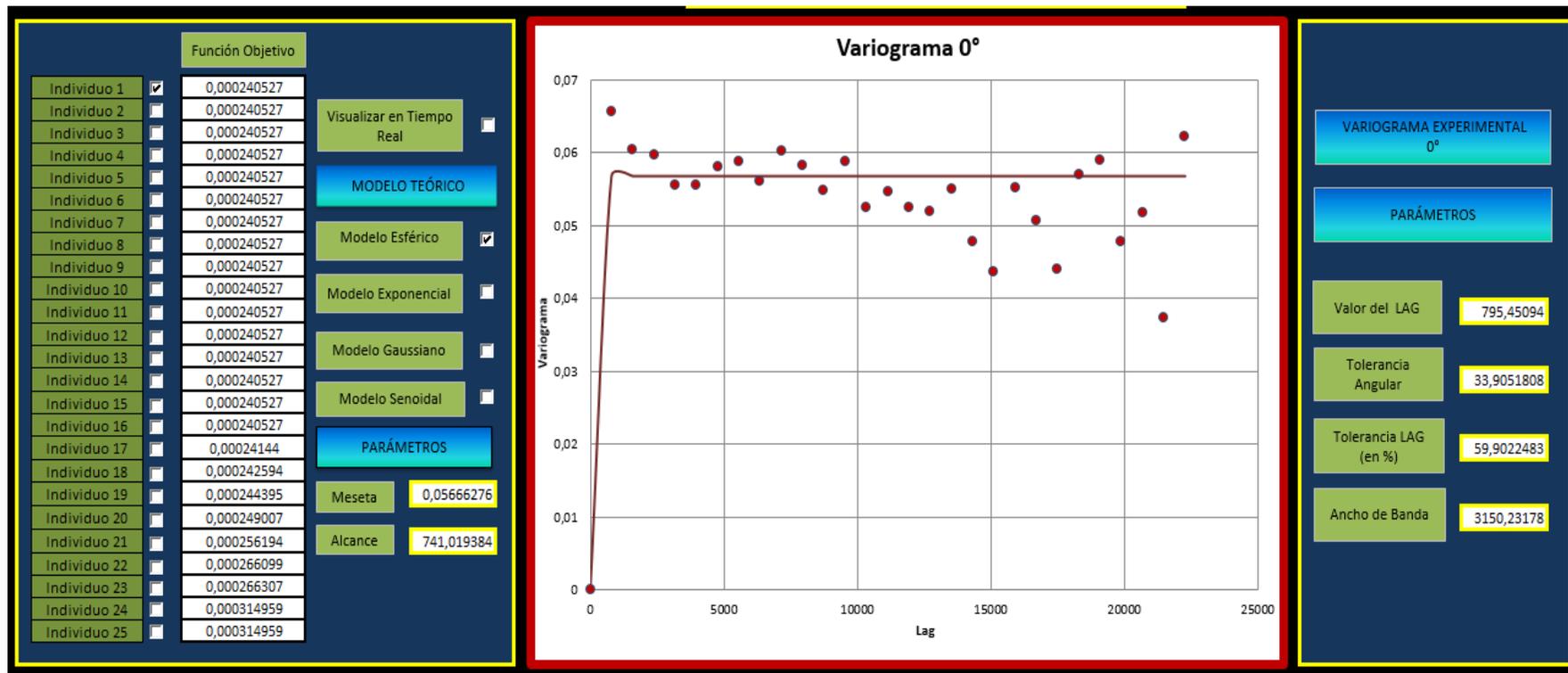


Figura 5.18 Mejor ajuste (Esférico) en el variograma 0° del campo “UPSE”

Elaborado por: Dennis Lucin

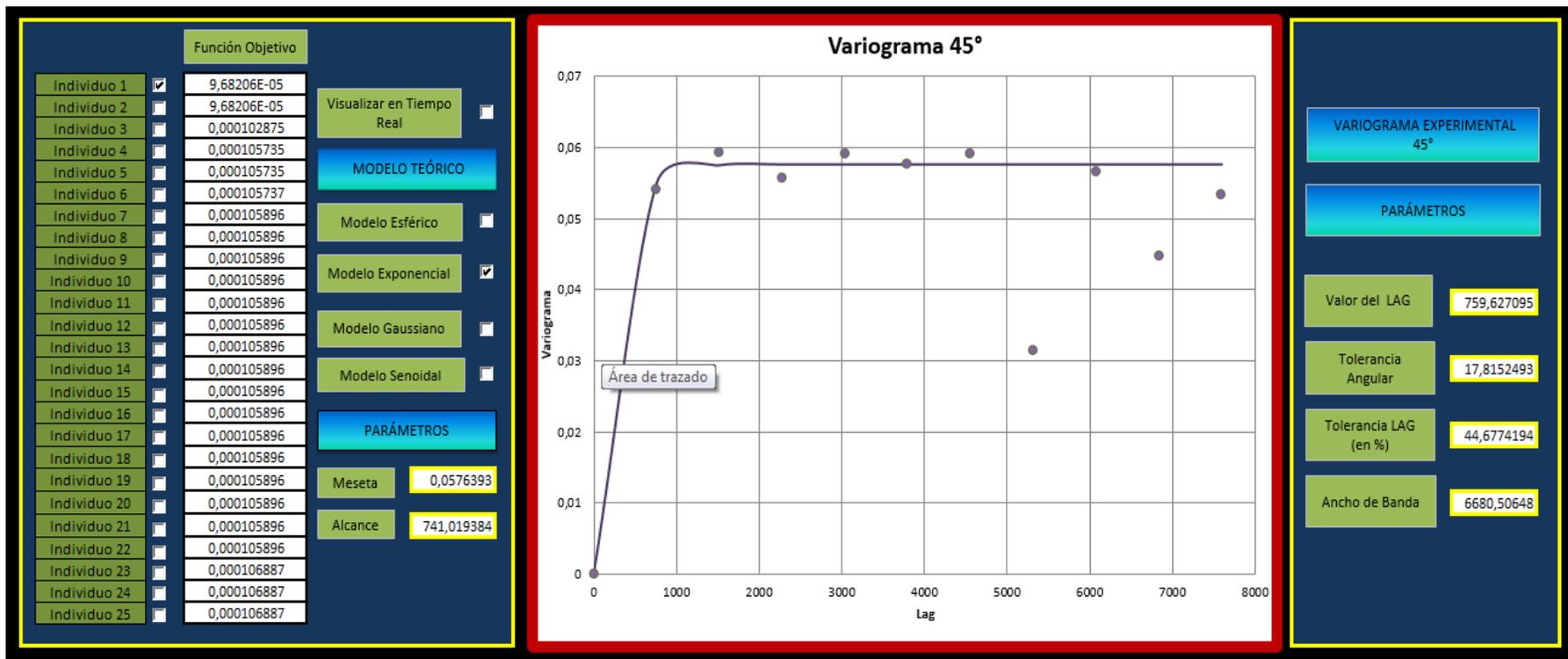


Figura 5.19 Mejor ajuste (Exponencial) en el variograma 45° del campo “UPSE”

Elaborado por: Dennis Lucin

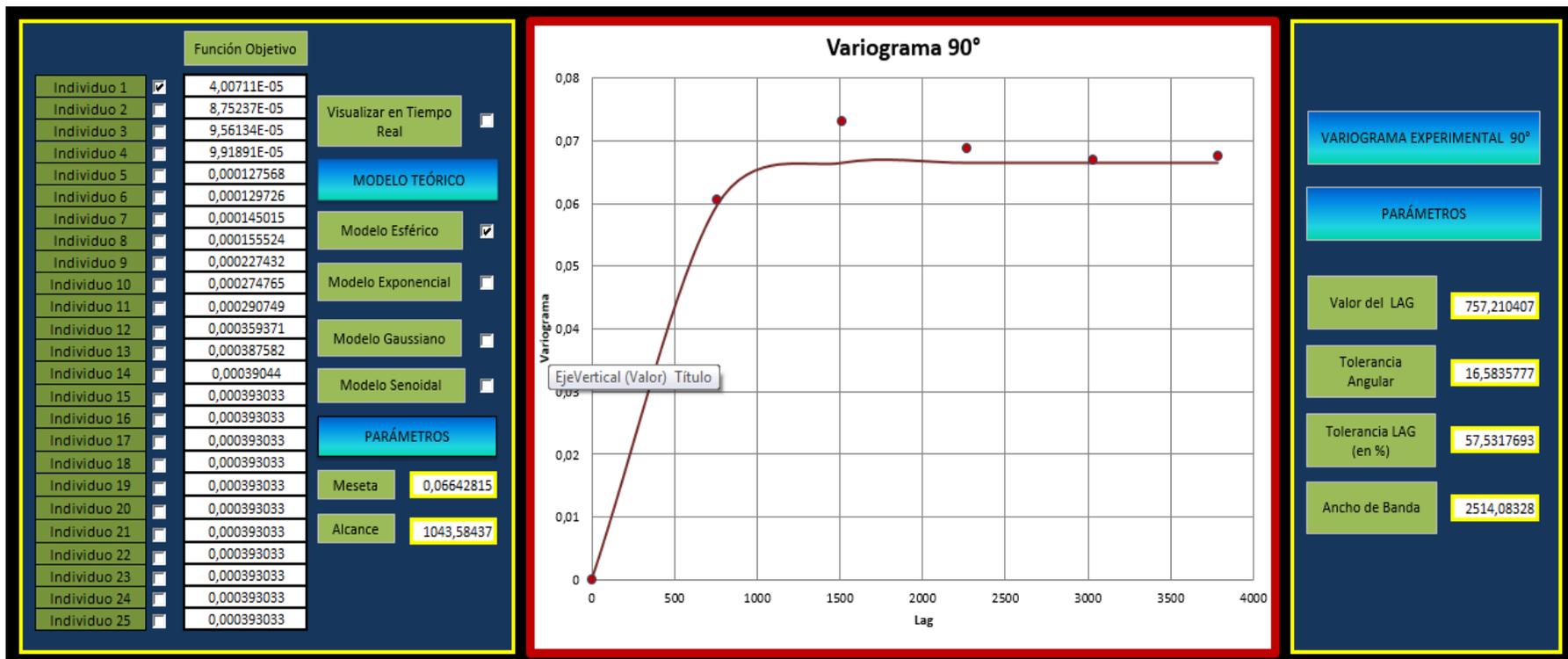


Figura 5.20 Mejor ajuste (Esférico) en el variograma 90° del campo “UPSE”

Elaborado por: Dennis Lucin

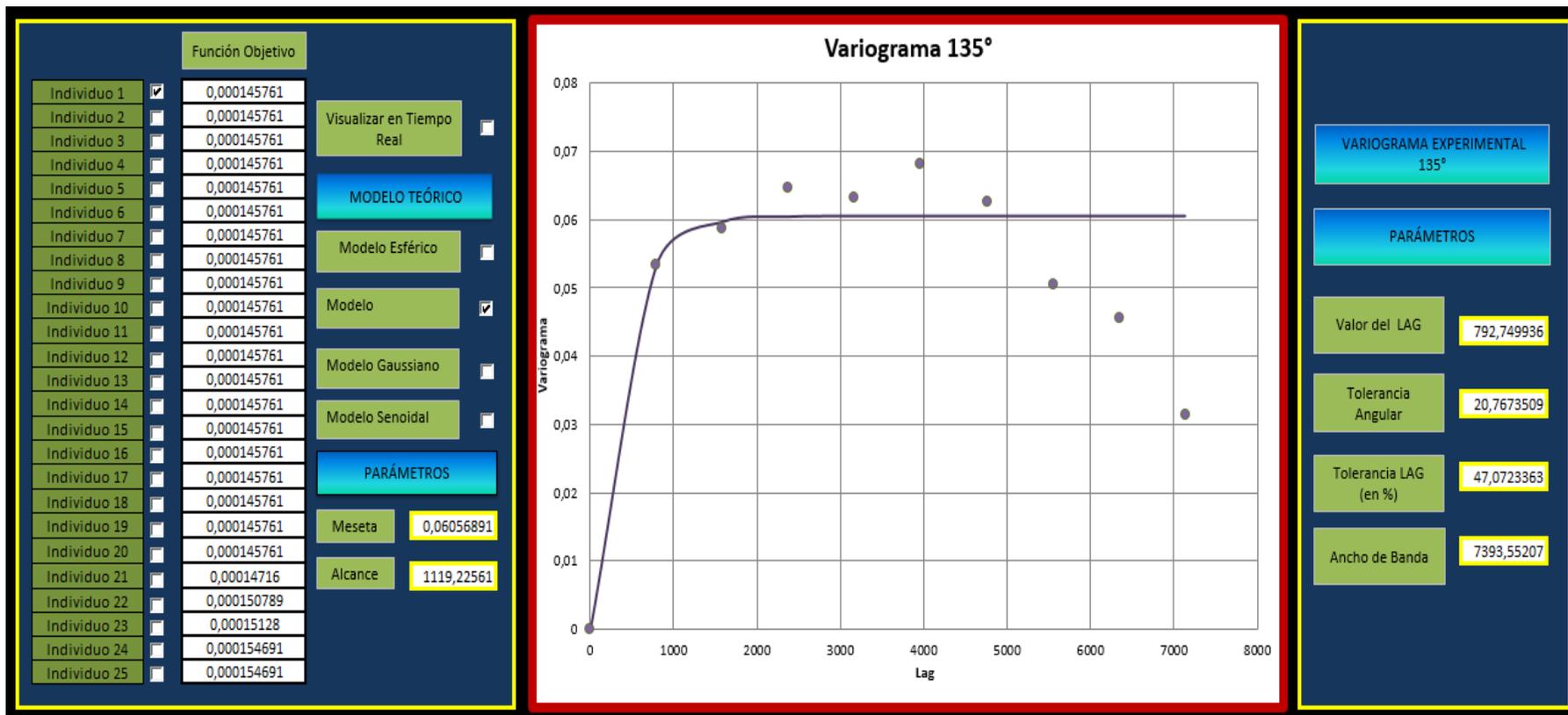


Figura 5.21 Mejor ajuste (Exponencial) en el variograma 0° del campo “UPSE”

Elaborado por: Dennis Lucin

5.5 RESULTADOS DE LAS ELIPSES DE ANISOTROPIAS DEL CAMPO “UPSE”.

A través del ajuste de estos variogramas se obtienen los valores como el alcance y meseta, valores que son necesarios para la construcción de las elipses de anisotropía del campo “UPSE”; a partir de ellas obtenemos parámetros como rango mayor, menor y dirección preferencial de la elipse. Las figuras 5.22 y 5.23 muestran las elipses de anisotropía construidas a partir del alcance resultante del ajuste de los variogramas, estas fueron realizadas por modelo Teórico.

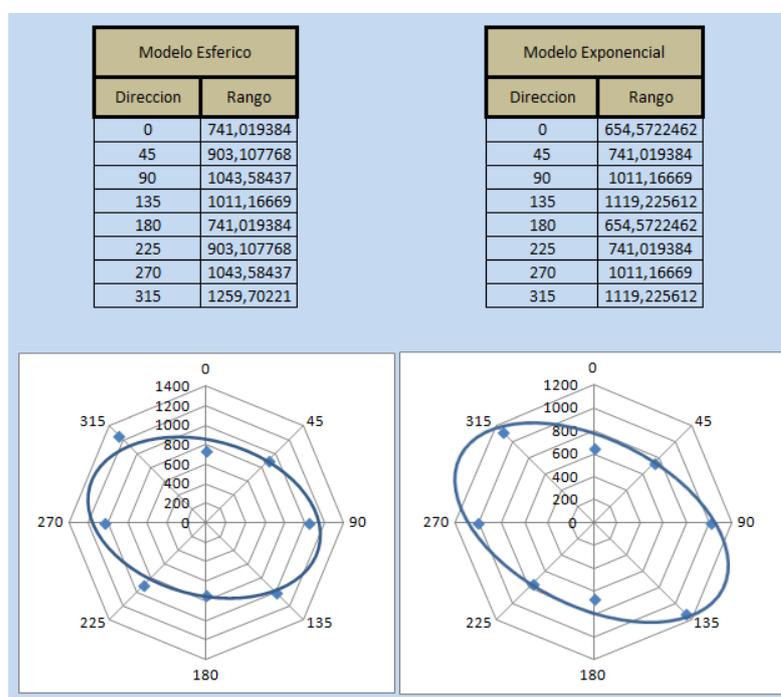


Figura 5.22 Elipses de anisotropía del modelo Esférico y Exponencial del campo “UPSE”.

Elaborado por: Dennis Lucin

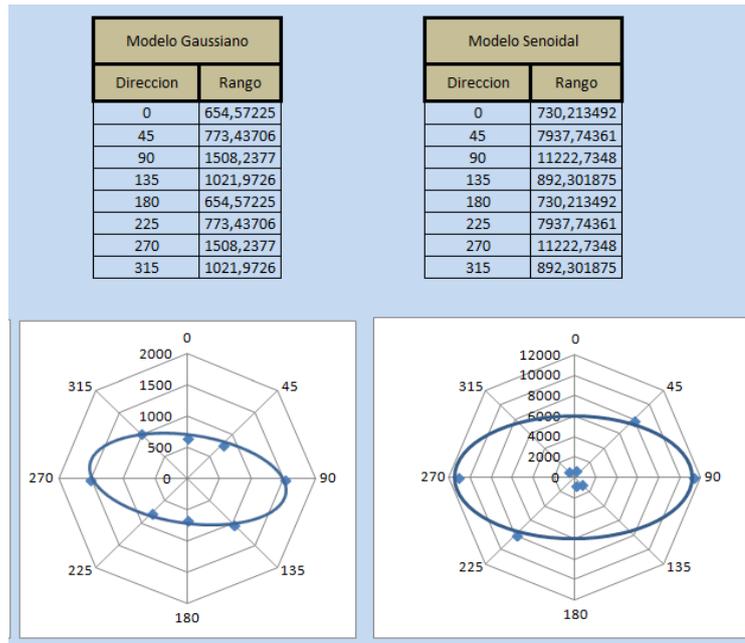


Figura 5.23 Elipses de anisotropía del modelo Gaussiano y Senoidal del campo “UPSE”.

Elaborado por: Dennis Lucin

Finalmente se obtuvo la elipse de anisotropía con los modelos que mejor se ajustaron a la dirección de 0°, 45°, 90°, y 135° (figura 5.24)

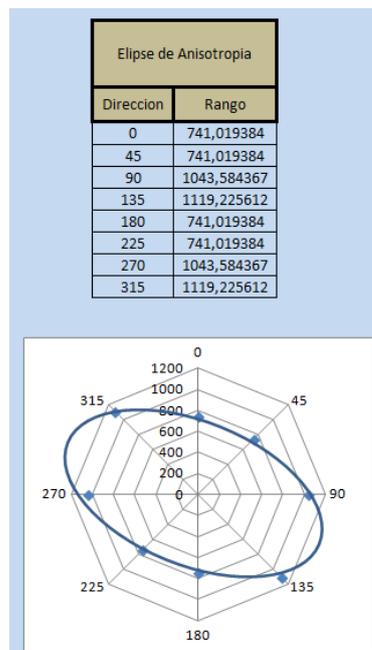


Figura 5.24 Elipses de anisotropía del mejor ajuste del campo “UPSE”.

Elaborado por: Dennis Lucin

De acuerdo a los resultados obtenidos de las elipses de anisotropía, se puede distinguir que la tendencia preferencial de la arena U del campo UPSE tiene una dirección Este-Oeste (252°). De esta forma concuerda lo mostrado por el algoritmo con el resultado del análisis sedimentológico sobre la tendencia depositacional (dirección) del campo, mencionado en la sección 5.4.3. La figura 5.25 además muestra visualmente la dirección en donde los valores son semejantes, tendencia que resulta en una dirección aproximadamente igual al resultado obtenido por el algoritmo.

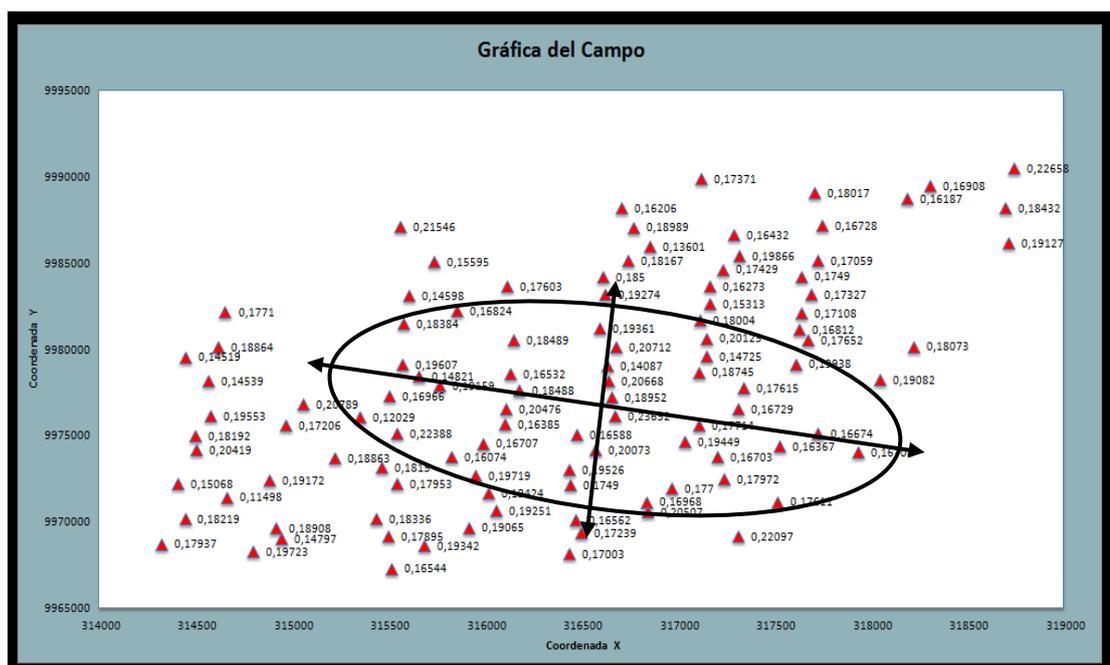


Figura 5.11 Dirección preferencial de la porosidad en la arenisca U del campo UPSE.

Elaborado por: Dennis Lucin

Cabe recalcar que para una mayor comparación del resultado, se debe verificar lo obtenido con los mapas de facies de la arena, lo cual no se contaba con dichos mapas del campo UPSE. Pero el análisis sedimentológico realizado a varios núcleos de la arena U, mostraba en general la tendencia depositacional, definido por el tipo de ambiente de depositación. De ésta forma y con lo observado en la distribución de la porosidad, es decir en qué dirección los valores son uniformes, se demostró

y comparó los resultados alcanzados tanto del análisis como lo logrado con el algoritmo.

Para éste mismo campo se realizó varias corridas del software con diferentes parámetros del algoritmo, dando como resultado elipses aproximadamente iguales, la mejor respuesta depende exclusivamente de que tan consistentes son los parámetros ingresados.

Si se realiza varias corridas para el mismo campo con diferentes parámetros de ingresos se obtiene que el ajuste variográfico varía lo que provoca una leve diferencia de resultados en las elipses de anisotropía y, generalizar una sola respuesta mediante la utilización de VariogramGenetics resulta algo difícil, ya que cada corrida dará un nuevo resultado que podría variar entre un 5 y 20% de la respuesta real, ya que la estimación de estos parámetros se realiza en un espacio de búsqueda muy grande, y al ser múltiples parámetros entonces la respuesta muy difícilmente llegará a ser la misma.

Capítulo VI

6 Conclusiones y Recomendaciones

6.1 CONCLUSIONES

A partir de la investigación pertinente y de las diversas corridas de VariogramGenetics (mencionadas y no mencionadas en este proyecto) se puede concluir que:

- Es posible la optimización de parámetros del cálculo de variogramas, mediante la programación de un Algoritmo Genético a través de Macros de Visual Basic en Microsoft Excel, con una función objetivo definida, de tal manera que estos parámetros van evolucionando y adaptándose a la función.
- El algoritmo genético ha probado ser de gran ayuda en la determinación del mejor ajuste variográfico, donde no se tiene conocimiento de cuál es la mejor combinación de valores que lleva a una solución óptima.
- Se consigue un equilibrio entre la eficacia y la eficiencia. Este equilibrio es configurable mediante los parámetros y operaciones usados en el algoritmo. Así, por ejemplo, bajando el valor de la tolerancia de la función objetivo conseguiremos una rápida solución a cambio de perder en "calidad", es decir, se obtendrá un ajuste variográfico no tan óptimo. Si aumentamos dicho valor, tendremos una mejor solución a cambio de un mayor tiempo consumido en la búsqueda. Por lo tanto, obtenemos una buena relación entre la calidad de la solución y el tiempo.
- El ajuste variográfico para el modelo teórico tipo senoidal es el que menos se ajusta con relación a la tolerancia. De ésta parte es la importancia de añadir una condición al algoritmo en caso de que no se llegue a una solución óptima. Se comprobó que en la mayoría de los casos para este modelo se necesita un valor de la tolerancia mayor con respecto a los demás modelos. Esto se debe a que en la

mayoría de los campos de prueba para el algoritmo, el variograma experimental no tenía una tendencia senoidal, razón por la cual no se ajustaba a la tolerancia ingresada.

- El resultado de la dirección de la elipse de anisotropía efectuado por el algoritmo genético, reflejó la misma dirección preferencial depositacional de la arenisca U que se detalló en el análisis sedimentológico del campo “UPSE”.
- Para el campo sintético se correlacionó los resultados obtenidos del algoritmo con la distribución de la propiedad, de tal forma que se comparó la tendencia de la porosidad con el resultado de las elipses de anisotropía, siendo ambos la misma dirección.

6.2 RECOMENDACIONES

Para que VariogramGenetics pueda realizar un mejor trabajo y para una mejor experiencia de usuario, es recomendable atenerse a las siguientes recomendaciones:

- Es de suma importancia la disposición de la información del campo a la que se realizará el análisis variográfico a través del algoritmo, es decir poseer cada dato de la propiedad y coordenadas por pozo perteneciente al campo. Esto permitirá que el resultado obtenido de las elipses de anisotropía sea efectivo.
- El ingreso de los parámetros del algoritmo se sugiere operar con los valores establecidos por el programa para el valor del lag, de ésta forma se asegura que se disponga de la mayor cantidad de lag para el cálculo de variograma experimental.

- Ejecutar el algoritmo con un porcentaje de mutación menor al 15% para poblaciones pequeñas y un 10% para poblaciones grandes
- Ejecutar VariogramGenetics en un equipo con mayor capacidad de memoria RAM, para que ésta funcione adecuadamente y con mayor rapidez.

REFERENCIAS BIBLIOGRÁFICAS

[1] Kelkar M., Pérez G., (2002). Geoestadística aplicada para la caracterización de reservorios. Spe.1-187, 226-238.

[2] Castaño A., Vergara F. Simulación geoestadística aplicada al modelamiento de Yacimientos de petróleo,

[3] Mateu J. e Morell I. Geoestadística y modelos matemáticos en hidrología,

[4] Giraldo R. Introducción a la Geoestadística teoría y aplicación.

[5] García F. Fundamentos de Geoestadística.

[6] Giraldo R. Estadística espacial,

[7] Ganan A., (2014). Caracterización de un reservorio aplicando técnicas geoestadísticas para un campo del distrito amazónico,

[8] Ginzo M. Análisis geostadístico de datos funcionales,

[9] Parra A. Adaptación de algoritmo de síntesis de texturas parasimulación geoestadística de múltiples puntos condicionada,

[10] Romero J., Dafonte C., Gómez A. y Penousal F., (2007). Inteligencia artificial y computación avanzada,

[11] Sontag E., (1972). Temas de inteligencia artificial.

[12] Hernández B. y Haces J. (2011). Tesis “sistema para generar micromundos para la asignatura de álgebra”.

[13] Mondavi D., (2014) Implementación de algoritmos genéticos para la optimización de parámetros de yacimiento mediante curvas de producción de pozos seleccionados en un campo del oriente ecuatoriano.

[14] Odorico A., Lage F. y Cataldi Z. Robótica, informática, inteligencia artificial y educación

[15] Ochoa G. Introducción a la computación evolutiva y la morfogénesis artificial,

[16] Melanie Mitchell (1998). An introduction to genetic algorithms,

[17] Arranz J. y Parra A. Algoritmos genéticos

[18] Rodríguez P. Introducción a los algoritmos genéticos y sus aplicaciones,

[19] http://books.google.com.ec/books?hl=es&lr=&id=kmjK7DJ_tJoC&oi=fnd&pg=PA15&dq=related:eBYR3RfeC2d0cM:scholar.google.com/&ots=aWZlyNRqBf&sig=ZAAIMuCArVVXW2nDSrPXXa4K-h4#v=onepage&q&f=false

[20] http://www.academia.edu/4232917/INTELIGENCIA_ARTIFICIAL_PRINCIPIOS_Y_APLICACIONESInformacion_del_Autor

[21] <http://www.ptolomeo.unam.mx:8080/jspui/bitstream/132.248.52.100/219/7/A7.pdf>

[22] <http://elcajondewatson.blogspot.com/2008/05/inteligencia-artificial-concepto-e.html>

[23] <http://www.actuarios.org/espa/revista12/17-art08.pdf>

[24] <http://www.dsi.fceia.unr.edu.ar/downloads/iaa/biblio/Que%20es%20la%20IA.pdf>

[25] http://www.ugr.es/~setchift/docs/conciencia_capitulo_2.pdf