

**DESARROLLO DE UN SIG-WEB CON CAPACIDAD DE PROCESAMIENTO
GEOESTADÍSTICO, APLICADO AL MODELAMIENTO DE LA VARIABILIDAD
ESPACIAL DE PROPIEDADES QUÍMICAS DEL SUELO EN EL SECTOR RÍO
FRAILE - RÍO AMAIME EN EL VALLE DEL CAUCA**



**CRISTIAN ANDRÉS MURILLO ALBORNOZ
JOSÉ MIGUEL VICTORIA URRESTI**

UNIVERSIDAD DEL VALLE
ESCUELA DE INGENIERIA CIVIL Y GEOMATICA
INGENIERIA TOPOGRÁFICA
Santiago de Cali, Agosto de 2011.

**DESARROLLO DE UN SIG-WEB CON CAPACIDAD DE PROCESAMIENTO
GEOESTADÍSTICO, APLICADO AL MODELAMIENTO DE LA VARIABILIDAD
ESPACIAL DE PROPIEDADES QUÍMICAS DEL SUELO EN EL SECTOR RÍO
FRAILE - RÍO AMAIME EN EL VALLE DEL CAUCA**

Presentado por:

**CRISTIAN ANDRES MURILLO ALBORNOZ
JOSÉ MIGUEL VICTORIA URRESTI**

Trabajo de Grado para optar por el título de
Ingeniero Topográfico.

DIRECTOR
JORGE ENRIQUE GARCÍA HURTADO
Ingeniero Agrícola.

UNIVERSIDAD DEL VALLE
ESCUELA DE INGENIERIA CIVIL Y GEOMATICA
INGENIERIA TOPOGRÁFICA
Santiago de Cali, Agosto de 2011.

NOTA DE ACEPTACIÓN

FIRMA PRESIDENTE DEL JURADO
DIRECTOR

FIRMA JURADO 1

FIRMA JURADO 2

Santiago de Cali, Agosto de 2011.

*A mis padres, por su incondicional apoyo.
A mis hermanas y a Luz Stella Mena por su acompañamiento.
A mis amigos y compañeros por su alegría y reconocimiento.*

Cristian Andrés Murillo Albornoz

*A mis padres y hermanos, en especial a mi madre por preguntar y alentarme día a día.
A toda mi familia por su entusiasmo.
A mis amigos y compañeros, por brindarme su ayuda en el momento preciso.*

José Miguel Victoria Urrestí.

AGRADECIMIENTOS

Agradecemos a todas las personas que de alguna manera se interesaron en este proyecto, que con su asesoría mostraron su incondicional apoyo, dicho acompañamiento sin duda nos permitirá cumplir con nuestra meta inmediata y proyectos a futuro.

¡Gracias!

*Cristian Andrés Murillo
José Miguel Victoria U.*

INDICE GENERAL

	Pág.
Resumen	1
1. Introducción.....	3
1.1 Contexto.....	3
1.2 Antecedentes.....	3
2. Planteamiento del Problema	6
2.1 Definición del problema	6
2.2 Descripción del problema	6
3. Justificación	7
4. Objetivos.....	8
4.1 Objetivo General.....	8
4.2 Objetivos Específicos	8
5. Marco conceptual	9
6. Marco Metodológico	23
6.1 Selección de la zona de estudio	23
6.1.1 Localización geográfica	23
6.1.2 Aspectos Descriptivos de la zona de estudio.	24
6.1.3 Descripción de los datos	24
6.2 Desarrollo del SIG.....	25
6.2.1 Fase de Identificación	25
6.2.1.1 Visión Global del Sistema.	25
6.2.1.2 Descripción de la Sistema.....	26
6.2.1.3 Alcances del sistema.....	26
6.2.1.4 Análisis de Requerimientos.....	27
Requerimientos no Funcionales del Sistema	27
Requerimientos Funcionales del Sistema	28
6.2.1.5 Descripción de Hardware.....	29
6.2.1.6 Selección de Herramientas	29
6.2.2 Fase de Diseño.....	31
6.2.2.1 Análisis Exploratorio de Datos.....	31
6.2.2.2 Análisis Estructural de Datos.....	31
6.2.2.3 Predicción Espacial.....	33
6.2.3 Fase de Implementación.....	34
6.2.3.1 Estructura de funcionamiento.....	34

6.2.3.2 Construcción de la base de datos	38
6.2.3.3 Construcción de la interfaz grafica	40
6.3 <i>Análisis geoestadístico</i>	47
6.3.1 Análisis exploratorio de datos	47
6.3.2 Análisis estructural de datos	48
6.3.3 Predicción espacial	49
7. Resultados	50
7.1 <i>Funcionamiento del aplicativo</i>	50
7.1.1 Acceso al aplicativo.....	50
7.1.2 Carga de datos al aplicativo.....	51
7.1.3 Lectura, actualización y borrado de datos.....	52
7.1.4 Uso de los módulos de análisis geoestadístico	52
7.1.4.1 Generación de gráficos y de estadísticas descriptivas	53
7.1.4.2 Generación y ajuste del semivariograma experimental	54
7.1.4.3 Funcionamiento módulo de Predicción Espacial.	55
7.1.5 Descarga de información geo-procesada.....	57
7.1.6 Validación de Información ingresada por el usuario	60
7.2 <i>Análisis Geoestadístico</i>	61
7.2.1 Análisis exploratorio de Datos.....	61
7.2.1.1 Análisis Grafico	62
7.2.2 Análisis Estructural de Datos	65
7.2.3 Predicción Espacial.....	67
7.2.4 Validación de resultados	68
8. Conclusiones	71
9. Recomendaciones	73
10. Bibliografía	74

INDICE DE TABLAS

Tabla 1. Libertades del usuario en el <i>software libre</i>	10
Tabla 2. Tipos de predictores kriging y sus propiedades.....	18
Tabla 3. Variables presentes en los puntos muestrales.	25
Tabla 4. Listado de Requerimientos no funcionales del sistema	28
Tabla 5. Listado de Requerimientos Funcionales del sistema.....	28
Tabla 6. Información espacial utilizada.	41
Tabla 7. Estadísticas Descriptivas profundidad 0- 30 cm.	61
Tabla 8. Estadísticas Descriptivas profundidad 30- 50 cm.	61
Tabla 9. Tipo de transformación empleada	64
Tabla 10. Estadísticas descriptivas con transformación. Profundidad 0 – 30 cm.	64
Tabla 11. Estadísticas descriptivas con transformación. Profundidad 30– 50 cm.....	64
Tabla 12. Parámetros de los semivariogramas teóricos.....	66
Tabla 13. Características del equipo 1.....	70
Tabla 14. Características del equipo 2.....	70

INDICE DE FIGURAS

Figura 1. Fotografía lote del cultivo de mango, área 5 ha (izquierda), Distribución de puntos de muestreo en lote del cultivo (derecha)	5
Figura 2. Función de densidad de probabilidad.	13
Figura 3. Tipos de autocorrelación, a) Autocorrelación espacial positiva. b) Autocorrelación espacial negativa. c) Ausencia.	15
Figura 4. Gráfica de función de semivariograma.	16
Figura 5. Comparación de los modelos exponencial, esférico y Gaussiano.....	19
Figura 6. Zona de estudio del proyecto y distribución espacial de los datos.....	23
Figura 7. Ajuste del semivariograma teórico.....	32
Figura 8. Semivariogramas teóricos soportados por el aplicativo web. a. Modelo Exponencial; b. Modelo Esférico. Modelo Gaussiano; d. Modelo Lineal.	33
Figura 9. Archivo muestra en formato csv.....	36
Figura 10. Esquema global del sistema. Arquitectura cliente – servidor	37
Figura 11. Arquitectura cliente - servidor.....	38
Figura 12. Función de procedimiento en R. para calculo y ajuste de un semivariograma Experimental. (parte1).....	39
Figura 13. Función de procedimiento en R. para calculo y ajuste de un semivariograma Experimental. (parte2).....	40
Figura 14. Estructura modular - Interfaz gráfica de usuario.	41
Figura 15. Módulo de acceso a SIG Web.	42
Figura 16. Módulo de registro a SIG Web.....	43
Figura 17. Herramientas del panel visor cartográfico.....	43
Figura 18. Panel de control de usuarios.	43
Figura 19. Panel de control de capas.	44
Figura 20. Panel leyenda.....	44
Figura 21. Módulo de análisis exploratorio.	45
Figura 22. Módulo de análisis estructural de datos.....	45
Figura 23. Módulo de predicción espacial – Kriging Lineales.	46
Figura 24. Módulo de ayuda.....	47
Figura 25. Creación de un nuevo usuario y mensaje de confirmación.....	50
Figura 26. Mensaje de error al ingresar al sistema.....	51
Figura 27. Campos requeridos para cargar un archivo y mensaje de confirmación.....	51
Figura 28. Asistente de edición de datos.....	52
Figura 29. Menú de selección de datos a analizar.....	52
Figura 30. Georreferenciación de datos cargados.....	53
Figura 31. Parámetros necesarios para el análisis gráfico y estadístico.....	53
Figura 32. Resultados del análisis estadístico.....	54
Figura 33. Gráficos estadísticos generados.....	54
Figura 34. Parámetros necesarios para la generación y ajuste	55
Figura 35. Semivariograma experimental (<i>izquierda</i>) y ajustado (<i>derecha</i>)	55
Figura 36. Selección del tipo de Kriging.....	56
Figura 37. Superficie raster sobre la variable acidez (pH). Vista sobre el visor cartográfico.....	56
Figura 38. Mapa de varianza de error.....	57
Figura 39. Asistente para descarga de datos.....	58

Figura 40. Red de muestreo (propiedades químicas de suelo sector río Fraile río Amaime).	59
Figura 41. Superficie raster, superficie de predicción sobre la variable pH profundidad 0-30cm.	59
Figura 42. Advertencia generada por el sistema.	60
Figura 43. Mensaje de error generado por el sistema.	60
Figura 44. Mensaje de error generado por el sistema.	60
Figura 45. Gráficos estadísticos generados mediante el SIG para la variable pH.	62
Figura 46. Semivariograma experimental (izquierda) y ajustado (derecha).	66
Figura 47. Comparación de superficies obtenidas para la variable Potasio.	69
Figura 48. Comparación de superficies obtenidas para la variable Magnesio.	69

INDICE DE ECUACIONES

Ecuación 1. Expresión matemática de interpolador Kriging.	15
Ecuación 2. Semivarianza.	16
Ecuación 3. Semivariograma.	16
Ecuación 4. Función modelo esférico.	18
Ecuación 5. Función modelo Exponencial.	19
Ecuación 6. Modelo Gaussiano.	19
Ecuación 7. Función modelo Efecto Pepita.	20
Ecuación 8. Función de modelo Nugget puro.	32
Ecuación 9. Función de modelo lineal, circular o esférico.	33
Ecuación 10. Función de modelos a.) Modelo Exponencial; b.) Modelo Gaussiano.	33

GLOSARIO

Autocorrelación espacial: relación entre un fenómeno y su correspondencia en un espacio geográfico.

Alcance: para el modelo esférico, la distancia en la que el modelo alcanza el valor máximo, o sill (meseta). Para los modelos gaussianos y exponencial, que se aproximan a la meseta asintóticamente. El modelo Nugget tiene una meseta con un rango de cero; el modelo lineal usa 'meseta/alcance' simplemente para definir la pendiente.

Anisotropía: en Geoestadística, es la situación cuando el semivariograma exhibe un comportamiento que varía con la dirección.

Calcio: elemento químico, de símbolo Ca.

Capacidad de Intercambio Catiónico: es la cantidad total de cationes que el suelo puede intercambiar por unidad de peso.

Capacidad de Intercambio Catiónico Efectiva: suma de Cationes Intercambiables de un suelo, incluyendo la Acidez titulable (Al + H).

Catión: es un ion (sea átomo o molécula) con carga eléctrica positiva, es decir, ha perdido electrones. Los cationes están presentes en el organismo en elementos químicos.

Variable aleatoria o estocástica: es una variable cuyos valores se obtienen de mediciones en algún tipo de experimento aleatorio.

Función Aleatoria: correlación de variables aleatorias que dependen de la posición.

Covarianza: es una medida estadística de la correlación entre dos variables.

Geoestadística: es una rama de la estadística que trata fenómenos espaciales, su interés primordial es la estimación, predicción, y simulación de dichos fenómenos.

Kriging: método de interpolación del valor medio ponderado donde los pesos asignados a las muestras minimizan la varianza del error, la que se calcula como una función del modelo de semivariograma y localizaciones de las muestras relacionadas unas con las otras, y del punto o bloque que está siendo estimado.

Magnesio: elemento químico de símbolo Mg.

Meseta (sill): límite superior de cualquier modelo de semivariograma acotado, al que tiende asintóticamente para grandes distancias. Para el modelo lineal, la relación 'sill/rango' se usa para definir la pendiente.

Modelo Esférico: es una de las funciones autorizadas que es usada frecuentemente como modelo de semivariograma, en combinación con el modelo Nugget.

Modelo Exponencial: es una de las funciones autorizadas que es usada frecuentemente como modelo de semivariograma, en combinación con el modelo Nugget.

Modelo Lineal: es una de las funciones autorizadas que es usada frecuentemente como modelo de semivariograma, en combinación con el modelo Nugget.

Modelo Nugget: modelo de varianza constante comúnmente usado en combinación con uno o más funciones cuando se ajustan modelos matemáticos a semivariogramas experimentales.

Potasio: elemento químico de símbolo es K.

Semi-Variograma: es sinónimo de 'variograma'. No hay acuerdo en la literatura geoestadística de cual término debe usarse y se usan ambos indistintamente.

Sodio: elemento químico de símbolo Na.

RESUMEN

Este proyecto tuvo como propósito principal el desarrollo de un Sistema de Información Geográfica (SIG) Web con capacidad de procesamiento geoestadístico y de dominio público. Se realizaron automatizaciones para cada etapa presente en un análisis geoestadístico a fin de determinar la dependencia espacial de una variable regionalizada y determinar su comportamiento mediante predicción espacial. Los datos utilizados para probar el funcionamiento del SIG corresponden a muestras de propiedades químicas de suelo en el sector río Fraile- río Amaime en el municipio de Palmira-Valle del Cauca, dichas muestras fueron georreferenciadas y almacenadas en una base de datos espacial, que a su vez interacciona con otras herramientas y tecnologías de código abierto. Un aspecto clave a tener en cuenta, es que los procedimientos geoestadísticos consolidados en el aplicativo SIG permitieron llevar a cabo un estudio integral sobre las propiedades tomadas en el sector río Fraile- río Amaime, como caso de estudio, caracterizando la distribución espacial de los niveles de concentración de dichas variables.

El aplicativo implementado cumplió con las etapas geoestadísticas: análisis exploratorio de datos, análisis estructural de datos, ajuste del semivariograma experimental a modelos teóricos y predicción espacial por diferentes tipos de Kriging¹ Lineales. El SIG se divide en módulos generales como: de ayuda, gestor de capas geográficas, carga de datos, lectura y edición de datos, visualización de resultados y descarga de datos producto del geoprocesamiento de la información espacial introducida por un usuario; al igual que los módulos específicos inherentes a las etapas en geoestadística.

Finalmente los resultados obtenidos con relación al caso de estudio indicaron que todas las variables presentaron estructuras de dependencia espacial, además la continuidad espacial en todas propiedades revelan isotropía geoestadística. Las superficies generadas permitieron establecer que la mayoría de las variables presentaban sus valores más altos en las zonas norte y centro de la zona de estudio, los más bajos por su parte se encontraban distribuidos en las zonas sur, oeste y nororiente.

Palabras Clave: Geoestadística, dependencia espacial, geoprocesamiento, software código abierto, SIG.

¹ Expresión no traducible, construida a partir del nombre de d.g. krige, geólogo y estadístico que propuso el método.

ABSTRACT

This project main purpose was the development of a Web based-Geographic Information System (Web-GIS), under a client-server approach capable to provide geostatistical processing tools and available to public access. Automations for geostatistical techniques were established in order to understand and determine the spatial dependence of analyzed variables. The used data to test Web-GIS functionalities were soil samples about chemical properties in the Fraile river- Amaime river zone located in Palmira - Valle del Cauca, these samples were georeferenced and stored in a spatial database, which interacts with other tools and open source technologies. A key aspect to consider is that the geostatistical procedures consolidated into the GIS application allowed to conduct a comprehensive study on the studied soil sample by characterizing the spatial distribution of its concentration levels.

The developed application allowed accomplishing with stages like exploratory analysis, structural analysis, adjustment of the experimental semivariogram to theoretical models and spatial prediction by different types of Kriging. GIS is divided into modules such as: support, manager of geographic layers, data loading, reading and editing data, results visualization and data download product geoprocessing spatial information entered by the user, and specific modules that are inherent to stages in geostatistics.

Achieved results after analysis suggested that all analyzed variables presented spatial dependency structures since their semivariances were not consistent with distance. Obtained surfaces allowed us to conclude that in one hand most of variables presented their higher values around north and center zones and in the other hand the lower values were distributed around south, east and north-east zones.

Keywords: Geostatistics, spatial dependence, geoprocessing, open source software, GIS.

1. INTRODUCCIÓN

1.1 CONTEXTO

Existe un considerable número de ejemplos en los que se ha utilizado la geoestadística en estudios ambientales con el ánimo de predecir y simular un fenómeno²; hidrología, geología, epidemiología, agronomía, ciencias de suelo, ingeniería forestal, monitoreo ambiental, procesamiento de imágenes de satélite, astronomía o cualquier otra disciplina que trate con datos colectados en diferentes localizaciones espaciales para desarrollar modelos que involucren la estructura de dependencia entre las observaciones.

Aunque los sistemas de información geográfica (SIG) han permitido a los investigadores el análisis espacial de información, la modelación de patrones de distribución espacial es poco factible si no se cuenta con un programa estadístico adecuado. La geoestadística es la herramienta fundamental para ese trabajo. La aplicación de la geoestadística puede responder al problema de estimar el valor de una variable regionalizada en punto en donde esta no ha sido medida, a partir de un conjunto de puntos con valor conocido y así conocer el comportamiento de dicha variable. Dicha técnica en conjunto con los SIG se convierte en una herramienta de gran utilidad para la comprensión del comportamiento de la variable analizada así como los procesos que lo influyen.

Partiendo de la integración entre técnicas geoestadísticas y SIG, el presente proyecto desarrolló un aplicativo SIG implementado con herramientas de *software* libre y de dominio público en Web con capacidad de procesamiento geoestadístico, con la finalidad de obtener superficies que caractericen el comportamiento de la variable regionalizada almacenada en el aplicativo. De igual forma como caso de estudio se determinó el comportamiento de las propiedades químicas del suelo en el sector río Fraile- río Amaime en el departamento del Valle del Cauca.

1.2 ANTECEDENTES

Los siguientes trabajos son considerados para la sección de análisis de este proyecto; sus metodologías se toman en cuenta para el caso de estudio, el cual es modelar la variabilidad espacial de propiedades químicas del suelo en zona de estudio.

² Fenómeno: manifestación que se hace presente a la conciencia de un sujeto y aparece como objeto de su percepción.

(Vásquez P. 2009) estudió la variabilidad espacial de algunas propiedades físicas y químicas del suelo utilizando geoestadística y enfoques invariados y multivariados en un conjunto de lotes destinados a la actividad agrícola en el departamento del Magdalena. Las propiedades físicas (humedad gravimétrica, textura, densidad aparente, resistencia a la penetración y conductividad hidráulica) y químicas (MO, pH, CIC, N, P, K, Ca, Mg, Na, Cu, Zn, Fe, Mn, B) fueron obtenidas utilizando una cuadrícula (29*29 metros) para un total de 184 puntos.

El estudio fue elaborado en tres partes, la primera de estas consistió en un análisis de los datos durante la cual se determinaba la variación de cada una de las muestras. Durante la segunda parte se llevó a cabo un análisis geoestadístico de los datos, en esta fase se visualizó y graficó la variabilidad espacial de las propiedades evaluadas. Finalmente, la tercera parte consistió en la realización de un análisis de clasificación cuyo objetivo fue el de definir zonas homogéneas con condiciones de suelo relativamente similares con el objetivo de facilitar la implementación de prácticas de manejo agronómico.

(Giraldo y otros. 2008) estudio la distribución espacial de las propiedades edáficas dentro de un lote de mango ubicado en la Mesa - Cundinamarca tomando un área de estudio de 5ha su objetivo principal fue relacionar las incidencias de dicha distribución con la variación de rendimiento del cultivo. La metodología se componía de las siguientes etapas: análisis exploratorio de datos, análisis estructural, e interpolación por el método de Kriging.

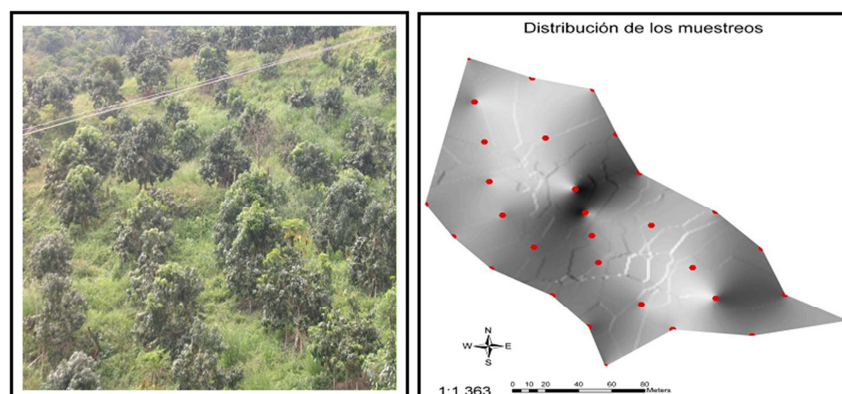
Para el análisis exploratorio de datos se realizó un muestreo materializando una red flexible en donde se ubicó 31 puntos georreferenciados con un dispositivo GPS, para cada punto se tomaron 5 sub-muestras de las cuales 4 estaban separadas a 1 m. de la sub-muestra central, resultando en total 151 muestras en total, el muestreo en el lote del cultivo consideró variables químicas en el suelo, además se midió para el rendimiento del cultivo contando el número de frutos por árbol en cada punto de muestreo, posteriormente se calcularon estadísticas descriptivas como: promedio desviación estándar, coeficiente de variación, valor máximo y mínimo en los datos medidos.

Para el análisis estructural se involucró dependencia espacial que se realizó para definir las relaciones entre variables. Además se analizó la tendencia espacial y el supuesto de normalidad (*Shapiro Wilk*³) para luego realizar semivariogramas experimentales y ajustarlos a modelos teóricos; para evaluar la calidad de estos modelos se utilizó el criterio de selección dado por los estadísticos de ajustes (*MAPE*⁴ y *CME*⁵) provenientes de la validación cruzada.

³ *Shapiro Wilk*, se usa para contrastar la normalidad de un conjunto de datos. Se plantea como hipótesis nula que una muestra $x_1, \dots, x_n$ proviene de una población normalmente distribuida.

⁴ *Mape*, Error absoluto porcentual de la media.

⁵ *CME*, Cuadrado medio del error.



**Figura 1. Fotografía lote del cultivo de mango, área 5 ha (izquierda),
Distribución de puntos de muestreo en lote del cultivo (derecha)**

Fuente: Documento tipo presentación (Giraldo y otros. 2008).

En la etapa de predicciones se realizó las interpolaciones espaciales al seleccionarse el modelo definitivo de semivarianza utilizando diferentes tipos de *Kriging* (*Logarítmico, residual y ordinario*)⁶. Se elaboró mapas de contorno además se realizó el análisis de componentes principales para determinar variables de mayor influencia en la fertilidad de suelo en estudio, reduciendo las variables al seleccionarse las que presentaron mayor número de relación significativas y las más importantes para el desarrollo y productividad del mango, establecidas mediante el coeficiente de correlación para generar áreas de manejo agrícola para el cultivo de mango.

(Acevedo, y otros. 2008) caracterizaron la variabilidad espacial de las propiedades químicas del suelo y su uso en el diseño de experimentos se analizaron propiedades en el suelo como: pH, materia orgánica, nitrógeno total, fósforo disponible y potasio intercambiable, utilizando el enfoque estadístico clásico (varianza y coeficiente de variación) y el de la geoestadística (semivarianza), para la obtención de valores representativos y mapas que muestren su comportamiento.

Se realizó un muestreo de suelos a una profundidad de 0-20 cm en cuadrícula, con equidistancias de 25 m y se obtuvieron 182 muestras en una parcela agrícola con una superficie de 12.7 ha. Con estos datos se analizó la variabilidad espacial desde el enfoque de la estadística clásica. Al aplicar el enfoque geoestadístico se encontró que todas las variables presentaron estructura espacial, manifestada en el semivariograma experimental, y se ajustaron a los modelos teóricos esférico, gaussiano lineal, lineal y efecto de agujero.

⁶ Véase capítulo MARCO CONCEPTUAL; [Kriging](#).

Según el enfoque de la estadística clásica, la mayoría de las variables estudiadas se sesgaron a la derecha, a excepción del nitrógeno total, que se ajustó a una función de probabilidad normal. Todas las variables estudiadas presentaron estructura espacial. Con los mapas generados por el método de Kriging fue posible identificar áreas parciales con diferente variabilidad, así como la dirección de mayor variabilidad de la propiedad en función de la distancia. Esto es importante para ubicar y orientar las parcelas experimentales.

2. PLANTEAMIENTO DEL PROBLEMA

2.1 DEFINICIÓN DEL PROBLEMA

Ausencia de una herramienta Web de fácil acceso y manejo que permita el procesamiento geoestadístico automatizado de variables cuantitativas y continuas en el espacio.

Desconocimiento del comportamiento de la variabilidad espacial de las propiedades químicas del suelo en el sector río Fraile - río Amaime.

2.2 DESCRIPCIÓN DEL PROBLEMA

En diversos proyectos en el campo de la geo-ciencias es frecuente estudiar variables que están espacialmente correlacionados. Generalmente, el estudio de dichas variables es realizado con *software* especialmente diseñado para dicho propósito, entre las que se encuentran las herramientas SIG. Tales herramientas, son distribuidas como aplicaciones de escritorio comercial y libre, las cuales comparten una limitante en común; la necesidad de encontrarse instaladas en el equipo del usuario, con las respectivas restricciones que esto conlleva en cuanto a plataformas, requerimientos de hardware y portabilidad. Por lo anterior es posible inferir que no existe una herramienta SIG que integre eficiencia y veracidad en sus resultados y que a su vez sea independiente del equipo y plataforma utilizada.

Por otra parte, después de realizar una búsqueda bibliográfica en entidades regionales encaminadas al estudio del suelo en el Departamento del Valle del Cauca y en redes académicas, se identificó que son nulos los estudios acerca del análisis de la variabilidad espacial de propiedades químicas de suelo en el sector río Fraile - río Amaime desde un enfoque geoestadístico, desaprovechando la capacidad de dichos estudios para aportar información que brinde soporte a la toma de decisiones y a la mejor gestión del recurso suelo.

3. JUSTIFICACIÓN

Los SIG no han sido ajenos a la creciente tendencia de compartir y procesar información mediante medios como la Web, actualmente vemos estas herramientas publicadas en Internet con diferentes propósitos, sin embargo en su mayoría se orientan a la visualización y consulta de información geográfica, a pesar de esto, los últimos avances tecnológicos han dado la posibilidad de incorporar a los SIG la capacidad de geoprocесamiento⁷ y generación de dicha información, eliminando la necesidad de utilizar herramientas instaladas en un equipo específico para cumplir con tales propósitos.

Tal geoprocесamiento se realiza para múltiples fines, por ejemplo, un estudio de estimación⁸ de una variable regionalizada de forma continua, a partir de toma de datos puntuales, lo cual conlleva un factor de incertidumbre asociada. Estos estudios actualmente se realizan usando herramientas, que integran SIG y técnicas geoestadísticas, no obstante, son abordados por personas o instituciones que poseen los recursos económicos para considerar la opción del *software propietario* o los conocimientos técnicos en el caso de un *software libre* con características complejas, lo cual condiciona el manejo y acceso a dichas herramientas. Una herramienta de dominio público permitiría a un usuario en particular la oportunidad de evaluar y estimar las características del fenómeno en la zona de estudio de interés. Además si la integración apunta a ofrecer una amplia difusión de técnicas de análisis de datos espacialmente correlacionados, se hace público el conocimiento, colocando a disposición de usuarios remotos el acceso a una base de datos espacial centralizada y un mecanismo de procesamiento prescindiendo de una infraestructura de *software/hardware* complejo.

Al mismo tiempo, el conocimiento de los recursos naturales constituye la base para la planificación del uso sostenido de todo espacio geoeconómico. Una herramienta que permita conocer el valor de una variable regionalizada de forma continua, resolvería cuestiones con respecto al comportamiento espacial de dicha variable de forma rápida y sencilla en cualquier sector o nivel de investigación de la ciencia aplicada, que actué como una herramienta de gestión para la administración de recursos no renovables que resultan estratégicos para la sociedad actual.

⁷ Véase capítulo MARCO CONCEPTUAL; Geoprocесamiento.

⁸ La palabra estimación es utilizada exclusivamente para inferir sobre parámetros fijos pero desconocidos;

4. OBJETIVOS

4.1 OBJETIVO GENERAL

Desarrollar un aplicativo web para modelar la variabilidad espacial de un fenómeno determinado mediante procesamiento geoestadístico.

Determinar el comportamiento de las propiedades químicas del suelo en el sector río Fraile- río Amaime.

4.2 OBJETIVOS ESPECÍFICOS

- Construir funciones que permitan llevar a cabo cada una de las etapas correspondientes a un análisis geoestadístico.
- Construir una base de datos espacial que permita el almacenamiento y gestión de la información a utilizar para el procesamiento geoestadístico.
- Generar gráficos, estadísticas e información geográfica tipo raster como resultado del procesamiento geoestadístico realizado a la información seleccionada por el usuario.
- Implementar un Sistema de Información Geográfica usando una plataforma libre que permita la interacción entre el usuario, los datos y los resultados de forma fácil y sencilla.
- Analizar la distribución espacial de las variables químicas para el caso de estudio.
- Proporcionar un manual de apoyo que permita llevar a cabo el procesamiento geoestadístico de datos georreferenciados.

5. MARCO CONCEPTUAL

Sistema de Información Geográfica (SIG): una definición clásica (Tomlin, 1990), es un elemento que permite analizar, presentar e interpretar hechos relativos a la superficie terrestre. El mismo autor argumenta, no obstante, que ésta es una definición muy amplia, y habitualmente se emplea otra más concreta. En palabras habituales, un SIG es un conjunto de *software* y *hardware* diseñado específicamente para la adquisición, mantenimiento, análisis y uso de datos cartográficos con el fin de resolver problemas complejos de planificación y gestión.

Geoprocesamiento: es una operación de un SIG para manipular los datos espaciales. Una operación de geoprocesamiento típica toma un conjunto de datos de entrada, realiza una operación en ese conjunto de datos, y devuelve el resultado de la operación como un conjunto de datos de salida. Las operaciones más comunes incluyen la superposición de información geográfica, interpolación, procesamiento topológico, álgebra de mapas, y conversión de datos. El geoprocesamiento permite por definición, la gestión y análisis de información para la toma de decisiones.

Georreferenciación: es un neologismo que refiere al posicionamiento con el que se define la localización de un objeto espacial (representado mediante punto, vector, área, volumen) en un sistema de coordenadas y datum determinado. Este proceso es utilizado frecuentemente en los Sistemas de Información Geográfica.

La georreferenciación, en primer lugar, posee una definición tecnocientífica, aplicada a la existencia de las cosas en un espacio físico, mediante el establecimiento de relaciones entre las imágenes de raster o vector sobre una proyección geográfica o sistema de coordenadas. Por ello la georreferenciación se convierte en el eje central para los modelados de datos realizados por los Sistemas de Información Geográfica.

Base de Datos Espacial: es un conjunto exhaustivo no redundante de datos u objetos espaciales estructurados organizados independientemente de su utilización y su implementación en máquinas accesibles en tiempo real y compatible con usuarios concurrentes con necesidad de información.

Lenguaje de Procedimiento: (*stored procedure*, en inglés), es un programa (o procedimiento) el cual es almacenado físicamente en una base de datos. Su implementación varía de un gestor de bases de datos a otro. La ventaja de un procedimiento almacenado es que al ser ejecutado, en respuesta a una petición de usuario, es ejecutado directamente en el motor de bases de datos, el cual usualmente corre en un servidor separado.

PL/pgSQL: (*Procedural Language/PostgreSQL Structured Query Language*, en inglés) es un lenguaje imperativo provisto por el gestor de base de datos PostgreSQL. Permite ejecutar comandos SQL mediante un lenguaje de sentencias imperativas y uso de funciones, dando mucho más control automático que las sentencias SQL básicas.

Desde *PL/pgSQL* se pueden realizar cálculos complejos y crear nuevos tipos de datos de usuario. Como un verdadero lenguaje de programación, dispone de estructuras de control repetitivas y condicionales, además de la posibilidad de creación de funciones que pueden ser llamadas en sentencias SQL normales o ejecutadas en eventos de tipo disparador (*trigger*).

Servicio Web: es un conjunto de protocolos y estándares que sirven para intercambiar datos entre aplicaciones. Distintas aplicaciones de *software* desarrolladas en lenguajes de programación diferentes, y ejecutadas sobre cualquier plataforma, pueden utilizar los servicios web para intercambiar datos en redes de ordenadores como Internet. La interoperabilidad se consigue mediante la adopción de estándares abiertos.

Software Libre: es la denominación del *software* que respeta la libertad de los usuarios sobre su producto adquirido y, por tanto, una vez obtenido puede ser usado, copiado, estudiado, modificado y redistribuido libremente. Según la **Free Software Foundation**⁹, el *software* libre se refiere a la libertad de los usuarios para ejecutar, copiar, distribuir, estudiar, modificar el *software* y distribuirlo modificado.

De acuerdo con tal definición, un *software* es "libre" cuando garantiza las siguientes libertades:

Tabla 1. Libertades del usuario en el *software* libre.

Libertad	Descripción
0	La libertad de usar el programa, con cualquier propósito.
1	La libertad de estudiar cómo funciona el programa y modificarlo, adaptándolo a tus necesidades.
2	La libertad de distribuir copias del programa, con lo cual puedes ayudar a tu prójimo.
3	La libertad de mejorar el programa y hacer públicas esas mejoras a los demás, de modo que toda la comunidad se beneficie.
Las libertades 1 y 3 requieren acceso al código fuente porque estudiar y modificar software sin su código fuente es muy poco viable.	

⁹ La Free Software Foundation (Fundación para el software libre) es una organización creada en octubre de 1985 por Richard Stallman y otros entusiastas del software libre con el propósito de difundir este movimiento.

Arquitectura Cliente - Servidor: esta arquitectura consiste básicamente en un programa “cliente” que realiza peticiones a otro programa “servidor” que le da una respuesta a su solicitud. En este modelo las aplicaciones se dividen de forma que el servidor contiene la parte que debe ser compartida por varios usuarios, y en el cliente permanece sólo lo particular de cada usuario.

Los clientes realizan funciones como:

- Manejo de la interfaz de usuario.
- Captura y validación de los datos de entrada.
- Conexiones con varios servidores a la vez.

Los servidores realizan funciones como:

- Recepción de las solicitudes de los clientes.
- Control de accesos concurrentes a bases de datos compartidas.
- Procesamiento de solicitudes y envío de respuestas al cliente.

Estructura Raster: su característica básica es la creación de una trama de celdas o píxeles, en la que cada una de ellas tiene una única propiedad espacial. La celda es su elemento básico. En una estructura Raster, lo que interesa es la propiedad espacial más que los límites exactos. Su tamaño determina la escala de la imagen que se está introduciendo, lógicamente cuanto más pequeña sea ésta, mayor será la escala de resolución que se obtenga.

Interpolación: técnica que predice los valores que tomará la variable dependiente de un conjunto determinado de localizaciones obteniéndose una especie de mapa probabilístico compuesto por celdas (véase estructura raster) del modo en que se distribuye la variable. Podemos clasificar los distintos métodos de interpolación según varios criterios (*Burrough, 1986*)

- Según los puntos considerados para el cálculo de valores. Algunos métodos consideran que todos los puntos de los que disponemos tienen influencia sobre el valor a calcular en una celda. Estos modelos se conocen como globales. En otros, denominados locales, solo se considera un conjunto restringido de estos. Este conjunto puede establecerse por medio de un umbral de distancia (todos los situados a una distancia menor que el umbral), de conteo (los “n” puntos más cercanos), o bien ambos.
- La selección de este conjunto de puntos más cercanos (los de más influencia) es un aspecto importante en el rendimiento de los métodos de interpolación de este tipo. Cuando se trabaja con un número de puntos elevado, se hace inviable el cálculo de las distancias entre todos esos puntos para seleccionar los más cercanos. El uso de índices espaciales y

otras estructuras semejantes se hace necesario para poder aplicar eficientemente estos métodos de interpolación sobre dichos conjuntos con tal número de puntos.

En realidad, un método global puede entenderse como uno local con un umbral infinito, no existiendo una dicotomía estricta entre ambas clases.

- Según su valor en los puntos de partida. En algunos métodos, denominados exactos, los valores asignados a las coordenadas correspondientes a los puntos de origen son exactamente los recogidos en dichos puntos. En los métodos aproximados, el valor en esas celdas es el que corresponde al mejor ajuste, y no ha de coincidir necesariamente con el valor original.
- Según la inclusión o no de elementos probabilísticos. Diferenciamos entre métodos estocásticos (aquellos que emplean elementos probabilísticos) y métodos determinísticos (aquellos que no los emplean).

Los métodos de interpolación más usuales a partir de un conjunto de puntos distribuidos irregularmente pueden reducirse en los siguientes: por vecindad, ponderación en función inversa de la distancia, superficies de tendencia y Kriging.

Geoestadística: es una rama de la estadística que trata fenómenos espaciales (*Journel & Huijbregts, 1978*). Su interés primordial es la estimación, predicción y simulación de dichos fenómenos (*Myers, 1987*). Es una técnica que ofrece una manera de describir la continuidad espacial, que es un rasgo distintivo esencial de muchos fenómenos naturales, y proporciona adaptaciones de las técnicas clásicas de regresión para tomar ventajas de esta continuidad (*Isaaks & Srivastava, 1989*).

La modelación espacial es la adición más reciente a la literatura estadística. Usualmente dicha modelación concierne con la predicción espacial, pero hay otras áreas importantes como la simulación y el diseño muestral (*Cressie, 1989*). Todo estudio geoestadístico tiene que llevarse a cabo en tres etapas:

1. Análisis exploratorio de los datos. Se estudian los datos muestrales sin tener en cuenta su distribución geográfica. Sería una etapa de aplicación de la estadística. Se comprueba la consistencia de los datos, eliminándose los erróneos e identificándose las distribuciones de las que provienen.
2. Análisis estructural. Estudio de la correlación entre puntos o continuidad espacial de la variable en el espacio. Se calcula el semivariograma, u otra función que explique la variabilidad espacial, y se ajusta al mismo un semivariograma teórico.

3. Predicciones. Estimaciones de la variable en los puntos no muestrales, considerando la estructura de correlación espacial seleccionada e integrando la información obtenida de forma directa en los puntos muestrales, así como la conseguida indirectamente en forma de tendencias conocidas. Este es un proceso que calcula un promedio 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.

Distribución normal: en estadística y probabilidad se llama distribución normal, distribución de Gauss o distribución gaussiana, a una de las distribuciones de probabilidad de variable continua que con más frecuencia aparece en fenómenos reales.

La gráfica de su función de densidad tiene una forma acampanada y es simétrica respecto de un determinado parámetro. Esta curva se conoce como campana de Gauss.

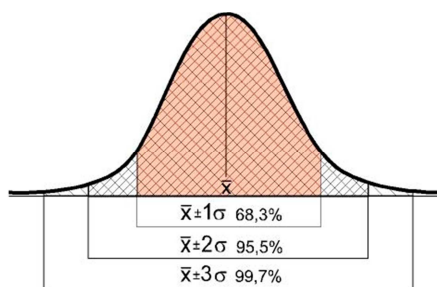


Figura 2. Función de densidad de probabilidad.

Fuente: Wikipedia.

La importancia de esta distribución radica en que permite modelar numerosos fenómenos naturales, sociales y psicológicos. Mientras que los mecanismos que subyacen a gran parte de este tipo de fenómenos son desconocidos, por la enorme cantidad de variables incontrolables que en ellos intervienen, el uso del modelo normal puede justificarse asumiendo que cada observación se obtiene como la suma de unas pocas causas independientes.

Histograma: En estadística, es una representación gráfica de una variable en forma de barras, donde la superficie de cada barra es proporcional a la frecuencia de los valores representados. En el eje vertical se representan las frecuencias, y en el eje horizontal los valores de las variables.

Diagrama de Cajas: un diagrama de caja es un gráfico, basado en cuartiles, mediante el cual se visualiza un conjunto de datos. Está compuesto por un rectángulo, la "caja", y dos brazos, los "bigotes".

Es un gráfico que suministra información sobre los valores mínimo y máximo, los cuartiles Q1, Q2 o mediana y Q3, y sobre la existencia de valores atípicos y la simetría de la distribución.

Diagrama de dispersión: un diagrama de dispersión es un tipo de diagrama matemático que utiliza las coordenadas cartesianas para mostrar los valores de dos variables para un conjunto de datos. Los datos se muestran como un conjunto de puntos, cada uno con el valor de una variable que determina la posición en el eje horizontal y el valor de la otra variable determinado por la posición en el eje vertical.

Gráfico Q-Q: en estadística, un gráfico Q-Q ("Q" viene de cuantil) es un método gráfico para el diagnóstico de diferencias entre la distribución de probabilidad de una población de la que se ha extraído una muestra aleatoria y una distribución usada para la comparación. Una forma básica de gráfico surge cuando la distribución para la comparación es una distribución teórica.¹ No obstante, puede usarse la misma idea para comparar las distribuciones inferidas directamente de dos conjuntos de observaciones, donde los tamaños de las muestras sean distintos.

Variable Regionalizada: se nombra como variable regionalizada $z(x)$ a la variable distribuida en el espacio de manera tal que presenta una estructura espacial de correlación.

Una definición más rigurosa matemáticamente equivalente consistiría en decir que una variable regionalizada es una variable aleatoria definida en un punto del espacio x . Donde en el caso más general x es un punto en el espacio tridimensional, es decir $x=(x_1, x_2, x_3)$.

Función Aleatoria: Si a cada punto x que pertenece a un dominio en el espacio le hacemos corresponder una variable aleatoria $Z(x)$, que en sentido general pueden ser dependientes, entonces el conjunto de variables aleatorias espacialmente distribuidas $\{Z(x), x, \in \Omega\}$ será una función aleatoria $Z(x)$.

Al tomar una muestra de una función aleatoria $Z(x)$, a la que llamaremos realización, se obtendrá una función espacial discreta $Z' = \{Z(X_i), X_i \in \Omega, i=1, \dots, n\}$ la cual constituye una *variable regionalizada*. Es decir una realización de una función aleatoria $Z(x)$ es una variable regionalizada Z' .

Variabilidad espacial: variación del comportamiento de una variable en un determinado entorno geográfico, esta es típicamente determinada mediante la predicción de valores en puntos no muestreados a través del uso de las interpolaciones. (Henríquez, 2005).

Autocorrelación espacial: es la concentración o dispersión de los valores de una variable en el espacio, es decir, a la existencia de una correlación de la variable consigo misma de tal modo que los valores de esta variable en un punto guardan relación directa con los de esta variable en otros puntos cercanos.

Si los valores de una variable son altos y suelen tener en su entorno valores también altos, y de modo similar sucede para valores bajos. Se dice que existe una autocorrelación espacial positiva. Puede, no obstante, existir una autocorrelación espacial negativa, si los valores altos se rodean de valores bajos y viceversa. En caso de no existir ningún tipo de autocorrelación espacial, se tiene que los datos recogidos en una serie de puntos son independientes entre sí y no se afectan mutuamente, con independencia de la distancia.

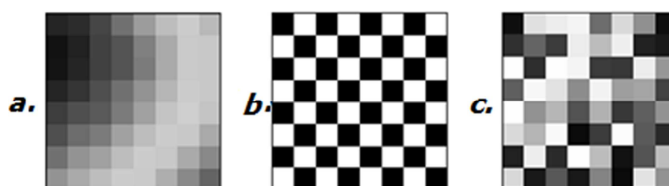


Figura 3. Tipos de autocorrelación, a) Autocorrelación espacial positiva. b) Autocorrelación espacial negativa. c) Ausencia.

Método Kriging: es un método de interpolación avanzado que genera una superficie estimada para una variable z a partir de un conjunto de puntos dispersos.

En su expresión fundamental, el Kriging es semejante a un método basado en ponderación por distancia. Dicha expresión es de la forma:

$$\hat{z} = \sum_{i=1}^n z_i \lambda_i; \quad \sum_{i=1}^n \lambda_i = 1$$

Ecuación 1. Expresión matemática de interpolador Kriging.

Siendo λ_i los pesos o coeficientes asignados a cada uno de los puntos considerados con la condición de que la suma de los coeficientes sea igual a 1 para que la estimación no tenga sesgo. El cálculo de estos pesos no obstante, se realiza de forma más compleja que en la ponderación por distancia ya que en lugar de utilizar dichas distancias se acude al análisis de autocorrelación a través del semivariograma. Para ello utiliza la semivarianza de las diferencias cuyo valor se calcula mediante la expresión:

$$\gamma(h) = \frac{1}{2} \cdot \text{var}(Z_p - Z_{p+h})$$

Ecuación 2. Semivarianza.

Donde **var** representa la varianza; Z_p el valor de la variable en un punto p y Z_{p+h} , el valor en un punto situado a una distancia h del anterior.

La función que relaciona γ con distancia h se denomina semivariograma y muestra por lo tanto la variación de la correlación entre los datos en función de la distancia. La expresión más usual para representarla es (Oliver y Webster, 1990):

$$\gamma(h) = \frac{1}{2 * N_h} \cdot \sum_{i=1}^{nh} (Z_{pi} - Z_{pi+h})^2$$

Ecuación 3. Semivariograma.

$N(h)$: Número total de pares de muestras de observaciones a una distancia h .

$Z(pi)$: Valor de muestra medida en punto i .

$Z(pi + h)$: valor de muestra medida en el punto $(pi + h)$.

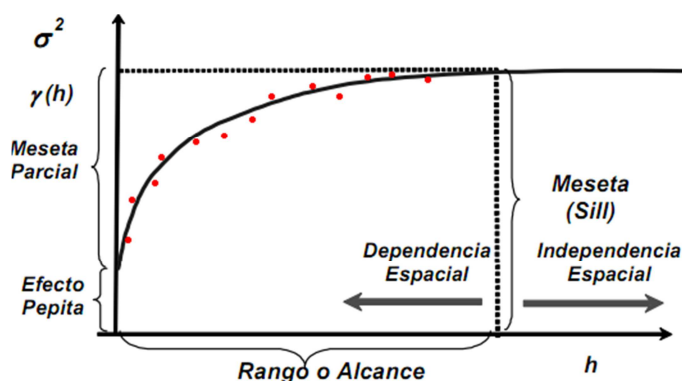


Figura 4. Gráfica de función de semivariograma.

Fuente: propia.

La forma del semivariograma está determinada por tres parámetros:

- Efecto pepita (*nugget*): que describe el ruido no correlacionado, y se denota por C_0 representando una discontinuidad puntual del semivariograma en el origen (Figura 4). Puede ser debido a errores de medición en la variable o a la escala de la misma. En algunas

ocasiones puede ser indicativo de que parte de la estructura espacial se concentra a distancias inferiores a las observadas.

- **Meseta (*sill*):** es la cota superior del semivariograma. También puede definirse como el límite del semivariograma cuando la distancia h tiende a infinito. La meseta puede ser o no finita. Los semivariogramas que tienen meseta finita cumplen con la hipótesis de estacionariedad fuerte; mientras que cuando ocurre lo contrario, el semivariograma define un fenómeno natural que cumple sólo con la hipótesis intrínseca. La meseta se denota por C_1 o por $(C_0 + C_1)$ cuando la pepita es diferente de cero. Si se interpreta la pepita como un error en las mediciones, esto explica porque se sugiere que en un modelo que explique bien la realidad, la pepita no debe representar más del 50% de la meseta. Si el ruido espacial en las mediciones explica en mayor proporción la variabilidad que la correlación del fenómeno, las predicciones que se obtengan pueden ser muy imprecisas. En la Figura 4 se representa este parámetro para el caso de uno de los modelos acotados.
- **Rango (*range*):** en términos prácticos corresponde a la distancia a partir de la cual dos observaciones son independientes. El rango se interpreta como la zona de influencia. Existen algunos modelos de semivariograma en los que no existe una distancia finita para la cual dos observaciones sean independientes; por ello se llama rango efectivo a la distancia para la cual el semivariograma alcanza el 95% de la meseta. Entre más pequeño sea el rango, más cerca se está del modelo de independencia espacial. El rango no siempre aparece de manera explícita en la fórmula del semivariograma. En el caso del modelo esférico (3.2.1), el rango coincide con el parámetro a , que se utilizará en las ecuaciones más adelante. Sin embargo, en el modelo exponencial (3.2.2), el rango efectivo es $a/3$ y en el modelo gaussiano (3.2.3) es $a/\sqrt{3}$.

El efecto pepita es la variabilidad espacial a distancias más pequeñas que la distancia de muestreo, más el error de medición (Rossi, 1992).

Para la determinación del semivariograma experimental se debe cumplir los siguientes principios fundamentales¹⁰.

1. El número de puntos muestrales para estimar un semivariograma en condiciones de isotropía debe ser 100. Si existe anisotropía los datos deben de ser mucho más abundantes y esto dependerá de las características de cada estudio.
2. El número de pares de datos necesarios en el cálculo de cada punto del semivariograma debe estar entre 30 y 50.

¹⁰ Isaals & Srivatava, 1989 y Goovaerts 1997.

3. Definido un semivariograma experimental, será necesario ajustarlo a un modelo teórico, debido a que el semivariograma empírico requiere de funciones matemáticas precisas. Dentro de los modelos teóricos se pueden encontrar el esférico, exponencial, gaussiano, potencial y lineal.

De acuerdo con (Olivier, 1987), (Cressie, 1993) y (Burrough y McDonnell, 1998), el semivariograma es la herramienta central de la geoestadística, ya que cuantifica la escala y la intensidad de la variación espacial además de proveer la información esencial para estimarla localmente utilizando el Kriging. Si no se tiene correlación espacial, entonces la semivarianza para cada par de puntos estará estimando la varianza total de los datos.

Kriging encierra un conjunto de métodos de predicción espacial que se fundamentan en la minimización del error cuadrático medio de predicción. En la tabla 2., se mencionan los tipos de Kriging y algunas de sus propiedades.

Tabla 2. Tipos de predictores kriging y sus propiedades.

Fuente: (Giraldo Henao, 2008).

Tipo de Predictor	Nombre	Propiedades
Lineal	- Simple - Ordinario - Universal - Indicador	- óptimos si hay normalidad multivariada. - Independiente de la Distribución son mejores predictores Linealmente insesgados
No Lineal	- Probabilístico - Log Normal, Trans-Gaussiano - Disyuntivo	Son predictores optimos

Modelo Esférico: tiene un crecimiento rápido cerca al origen (Figura 5.), pero los incrementos marginales van decreciendo para distancias grandes, hasta que para distancias superiores al rango los incrementos son nulos. Su expresión matemática es la siguiente:

$$\gamma(h) = \begin{cases} C_0 + C_1 \left(\frac{3}{2} \left(\frac{h}{a} \right) - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right) & h \leq a \\ C_0 + C_1 & h > a \end{cases}$$

Ecuación 4. Función modelo esférico.

En donde **C1** representa la meseta, **a** el rango y **h** la distancia.

Modelo Exponencial: este modelo se aplica cuando la dependencia espacial tiene un crecimiento exponencial respecto a la distancia entre las observaciones. El valor del rango es igual a la distancia para la cual el semivariograma toma un

valor igual al 95% de la meseta (Figura 5.). Este modelo es ampliamente usado. Su expresión matemática es la siguiente:

$$\gamma(h) = C_0 + C_1 \left(1 - \exp\left(\frac{-3h}{a}\right) \right)$$

Ecuación 5. Función modelo Exponencial.

Modelo Gaussiano: al igual que en el modelo exponencial, la dependencia espacial se desvanece solo en una distancia que tiende a infinito. El principal distintivo de este modelo es su forma parabólica cerca al origen Figura 5. Su expresión matemática es:

$$\gamma(h) = C_0 + C_1 \left(1 - \exp\left(\frac{-h^2}{a^2}\right) \right)$$

Ecuación 6. Modelo Gaussiano.

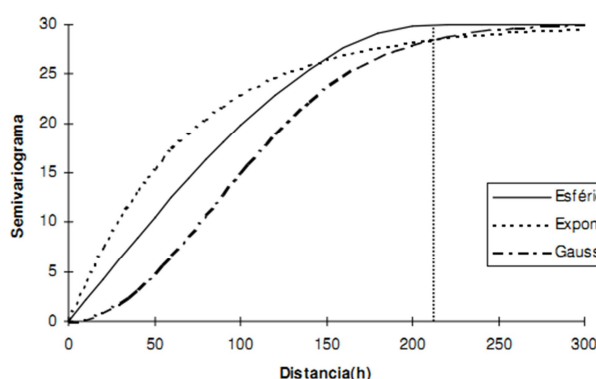


Figura 5. Comparación de los modelos exponencial, esférico y Gaussiano

Modelo de Independencia (Pepita Puro): es indicativo de carencia de correlación espacial entre las observaciones de una variable. Es común sumar este modelo a otro modelo teórico de semivarianza, para obtener lo que se conoce como semivariograma anidado. Lo anterior se sustenta en una propiedad de los semivariogramas que dice que cualquier combinación lineal de semivariogramas con coeficientes positivos es un semivariograma. Su expresión matemática es:

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ C_0 & h > 0 \end{cases}, \text{ donde } C_0 > 0$$

Ecuación 7. Función modelo Efecto Pepita.

Isotropía: esta característica indica si el proceso espacial que se está analizando cambia mucho según la dirección en la que se analicen los datos; el proceso se denomina como isotrópico en el caso en que la dirección en que se realice el análisis no sea un factor que condicione la forma de la distribución y por lo tanto al proceso espacial que la caracteriza; si se llega a encontrar anisotropía, esta es un indicador de que existen una serie de circunstancias que condicionan la forma de la distribución, Lo que puede entenderse como la integración entre diferentes procesos espaciales.

Suelo: es la cubierta superficial de la mayoría de la superficie continental de la Tierra. Es un agregado de minerales no consolidados y de partículas orgánicas producidas por la acción combinada del viento, el agua y los procesos de desintegración orgánica.

Los suelos cambian mucho de un lugar a otro. La composición química y la estructura física del suelo en un lugar dado, están determinadas por el tipo de material geológico del que se origina, por la cubierta vegetal, por la cantidad de tiempo en que ha actuado la meteorización, por la topografía y por los cambios artificiales resultantes de las actividades humanas.

El suelo como un sistema dinámico

El suelo es un cuerpo natural y dinámico que desempeña muchos roles claves en los ecosistemas terrestres (*Doran & Parkin, 1994*) y se forma a partir del material parental, el relieve, la biota, el clima y el tiempo, en una serie de procesos globales: ganancias, pérdidas, traslocaciones y transformaciones, que determinan sus características. La acción combinada de factores y procesos formadores conlleva al desarrollo de gran diversidad de suelos (*Buol, 1983*). Igualmente el suelo es el intermediario entre el sistema físico-climático (física dinámica de la atmósfera) y el sistema biogeoquímico (en la biosfera), por lo que cualquier actividad sobre él influye en procesos de nivel global. De acuerdo con la Teoría de Sistemas, el “sistema suelo” puede entenderse como un punto relativamente estable en la superficie terrestre frente a procesos de transferencia y transformación y, desde esta perspectiva, es evidente la relación morfología-procesos.

Edafología: ciencia que estudia las características de los suelos, su formación y su evolución (edafogénesis), sus propiedades físicas, morfológicas, químicas y mineralógicas y su distribución. También comprende el estudio de las aptitudes de los suelos para la explotación agraria o forestal.

pH: es una medida de la concentración de hidrógeno expresado en términos logarítmicos. El pH del suelo es generalmente considerado adecuado en agricultura si se encuentra entre 6 y 7. En algunos suelos, incluso con un pH natural de 8, pueden obtenerse buenos rendimientos agropecuarios. Sin embargo, a partir de tal umbral las producciones de los cultivos pueden mermarse ostensiblemente. En la mayoría de los casos, los pH altos son indicadores de la presencia de sales solubles, por lo que se requeriría acudir al uso de cultivos adaptados a los ambientes salinos.

Calcio: (elemento químico, de símbolo Ca), es un metal alcalinotérreo que interfiere con la absorción de K y NH₄, puede contribuir a deficiencias de magnesio y Boro. Un exceso de Ca generalmente origina un aumento del pH que precipita muchos micronutrientes, creando carencias de los mismos.

Sodio: (elemento químico, de símbolo Na). Es un metal blando, ligero y de color plateado que no se encuentra libre en la naturaleza su presencia produce suelos lodosos, con poca aireación y poca disponibilidad de agua por aumento de presión osmótica. No es un nutriente para las plantas, por lo tanto no es necesario para el crecimiento de las plantas. Altos niveles de sodio son perjudiciales para el suelo y el crecimiento de las plantas.

Potasio: (elemento químico, de símbolo es K). Es el quinto metal más ligero y liviano. Al igual que otros metales alcalinos reacciona violentamente con el agua desprendiendo hidrógeno, incluso puede inflamarse espontáneamente en presencia de agua. Puede causar deficiencias en Zn o Fe. El exceso contribuye a las deficiencias de nitrógeno, calcio y magnesio, zinc o hierro.

Magnesio: (elemento químico, de símbolo Mg). Este elemento no se encuentra en la naturaleza en estado libre (como metal), sino que forma parte de numerosos compuestos, en su mayoría óxidos y sales; es insoluble. Los niveles elevados de Magnesio suelen producir deficiencia de Calcio.

Capacidad de Intercambio Catiónico: es la cantidad total de cationes que el suelo puede intercambiar por unidad de peso. Es una medida de la capacidad del suelo para retener y liberar elementos tales como K, Ca, Mg y Na. Suelos con alta arcilla o Materia Orgánica, tienden a tener una alta CIC. Al CIC también se lo interpreta como la Fertilidad del suelo.

Capacidad de Intercambio Catiónico Efectiva: suma de Cationes Intercambiables de un suelo, incluyendo la Acidez titulable (Al + H). Difiere de la verdadera Capacidad de Intercambio Catiónico (CIC) en que esta evalúa el número total de puestos de carga negativa y la CICE evalúa solamente los puestos que están ocupados.

Kriging simple: es un tipo de Kriging el cual parte de la hipótesis básica de la estacionariedad junto con el hecho de que se asume que la media m , de la función aleatoria es conocida. Esto es,

$$E(Z(u)) = m \quad \text{y } m \text{ es conocida}$$

Kriging Ordinario: Generalmente el valor de la media m es desconocido y por lo tanto no se puede utilizar el Kriging simple. El Kriging ordinario establece una condición adicional al sistema de ecuaciones del Kriging simple para filtrar el valor desconocido de la media.

Al igual que antes el estimador propuesto es de la forma

$$Z^*(u) = \sum_{\alpha=1}^N \lambda_{\alpha}(u) Z(u_{\alpha})$$

Kriging Universal: Este tipo de Kriging permite tratar el caso que exista una tendencia. De esta tendencia se debe conocer su expresión matemática, es decir las ecuaciones que la describen y sólo ignorar los coeficientes que le asignan valores numéricos. En muchos casos, la variable no satisface estas condiciones y se caracteriza por exhibir una tendencia. Por ejemplo en hidrología los niveles piezométricos de una acuífero pueden mostrar una pendiente global en la dirección del flujo. Para tratar este tipo de variables es frecuente descomponer la variable $Z(x)$ como la suma de la tendencia, tratada como una función determinística, más una componente estocástica estacionaria de media cero. Asuma que:

$$Z(x) = m(x) + \varepsilon(x)$$

Con $E(\varepsilon(x)) = 0$, $V(\varepsilon(x)) = \sigma^2$, por consiguiente $E(Z(x)) = m(x)$.

La tendencia puede expresarse mediante:

$$m(x) = \sum_{i=1}^p a_i f_i(x)$$

Donde las funciones $f_i(x)$ son conocidas y p es el número de términos empleados para ajustar $m(x)$.

El predictor Kriging universal se define como:

$$Z^*(x_0) = \sum_{i=1}^n \lambda_i Z(x_i)$$

6. MARCO METODOLÓGICO

6.1 SELECCIÓN DE LA ZONA DE ESTUDIO

Por disponibilidad de datos en relación al levantamiento información por muestreo, para el presente proyecto se seleccionó como zona de estudio el sector río Fraile - río Amaime en el municipio de Palmira, Valle del Cauca; las características más relevantes, localización geográfica y los aspectos cualitativos de ésta se describen a continuación:

6.1.1 Localización geográfica

La zona está ubicada al sur del Departamento del Valle del Cauca entre la Cordillera Central y la margen derecha del Río Cauca, en jurisdicción del Municipio de Palmira. El área de estudio está limitada al sur, por los Ríos Fraile, Bolo, y Zanjón Chontaduro; al Norte con el Río Amaime; al Este con la carretera troncal de Occidente y al Oeste con el Río Cauca.

La zona de estudio está comprendida entre las coordenadas N-873.300 N-890.500 y E-1.065.800 a E-1.084.700, que cubre una extensión aproximada de 25.700 hectáreas. Dicha extensión depende de la distribución de puntos (muestras) tomados en trabajo de campo.

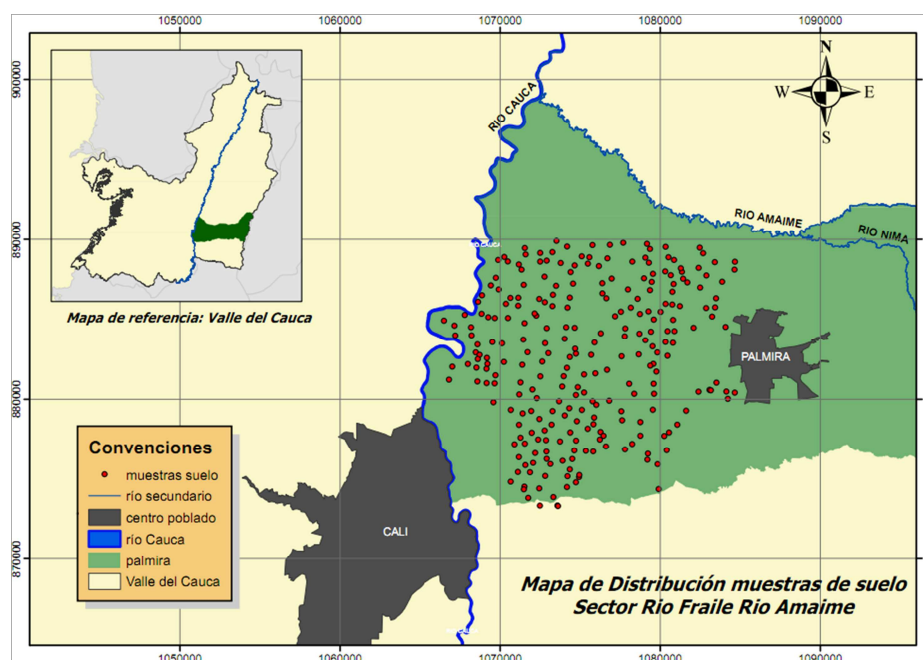


Figura 6. Zona de estudio del proyecto y distribución espacial de los datos.

6.1.2 Aspectos Descriptivos de la zona de estudio.

Geología

La zona de estudio se encuentra en el valle geográfico que está dominado por la formación valle o piso del valle, formado por sedimentos de acumulación fluvial.

Geomorfología

El área del sector Fraile – Amaime, está conformada por suelos en su mayoría de origen aluvial, debido a los cambios del lecho del Río Cauca y sus afluentes y los aportes de materiales conducidos por las corrientes que confluyen a la zona.

Topografía

Los suelos del Sector Fraile – Amaime, presentan en términos generales una topografía plana a ligeramente ondulada. La mayor pendiente se encuentra en el sentido Este – Oeste, con valores promedio de 0.4%. En el sentido Sur – Norte, la pendiente general es menor y tiene valores promedios de 0.03%.

6.1.3 Descripción de los datos

La información utilizada corresponde a un conjunto de datos de una jornada de muestreo de suelo del año 2008 realizada por la Corporación Autónoma Regional del Valle del Cauca (C.V.C) en la zona de estudio. Los datos recolectados estaban compuestos aproximadamente por 420 unidades de observación denominadas calicatas¹¹, esta información clasifico las calicatas según la taxonomía, dicho muestreo se realizó de forma irregular y a diferentes profundidades (entre 0 y 100 cm.).

Posteriormente estas fueron analizadas en laboratorio y se determinaron diferentes niveles de concentración en elementos químicos presentes en el suelo. La información general acerca del levantamiento de datos como los valores cuantitativos de los niveles de concentración en las muestras se describe en el **Anexo A** y **Anexo B** respectivamente.

La información mencionada anteriormente se encontraba recopilada en hojas de cálculo, de estas fueron descartados los datos que no tenían información acerca de su ubicación geográfica o que sus valores correspondientes a las propiedades químicas se encontraban incompletos o con incongruencias, una vez realizada la inspección, se obtuvieron un total de 262 puntos de muestreo, los cuales fueron utilizados como información base para las etapas del análisis geoestadístico. De acuerdo a la información de coordenadas (norte–este) en los puntos, posteriormente se georreferenció los datos asociando el sistema de proyección Bogotá 1975 / Colombia West Zone (EPSG¹²: 21891).

¹¹ Calicata; perforación que se practica para determinar la existencia de minerales o naturaleza del suelo.

¹² EPSG: (European Petroleum Survey Group). Base de datos sobre códigos que definen parámetros como: elipsoides, datums, sistemas de coordenadas, proyecciones cartográficas. <http://spatialreference.org/ref/epsg/>

Las variables seleccionadas para la realización del análisis geoestadístico fueron: calcio, sodio, magnesio, potasio, pH, capacidad de intercambio catiónico (C.I.Ca) y capacidad de intercambio catiónico efectiva (C.I.C.E). El estudio edafológico considero el muestreo a diferentes profundidades no obstante para este proyecto únicamente se tuvo en cuenta los dos primeros intervalos los cuales se definieron entre 0 a 30 y 30 a 50 centímetros.

Tabla 3. Variables presentes en los puntos muestrales.

Variable	símbolo	Unidad de medida¹³
pH (Acidez)	pH	-----
Calcio	Ca	me/100 g.s
Sodio	Na	me/100 g.s
Potasio	K	me/100 g.s
Magnesio	Mg	me/100 g.s
Capacidad de intercambio Catiónico	C.I.Ca	me/100 g.s
Capacidad de intercambio catiónico efectiva	C.I.C.E	me/100 g.s

6.2 DESARROLLO DEL SIG

En el presente seccion se definen cada una de las fases tenidas en cuenta para el desarrollo metodológico y puesta en marcha de la herramienta SIG.

6.2.1 Fase de Identificación

El propósito de esta fase fue proporcionar la información necesaria para llevar a cabo el desarrollo del SIG. Es decir, la visión del sistema, el alcance, definición de requerimientos (funcionales y no funcionales) y selección de herramientas.

6.2.1.1 Visión Global del Sistema.

El Sistema de Información Geográfica en la Web será una herramienta novedosa tecnológicamente y de proyección social que aportará conocimientos acerca de comportamiento un fenómeno natural estudiado.

¹³ Unidad de medida: mili-equivalentes por 100 gramos suelo.

En este orden, el SIG se encargara de almacenar, manipular y procesar información georreferenciada acerca de puntos de muestreo siguiendo algunos criterios geoestadísticos, con el objetivo de analizar y predecir la variabilidad espacial de un fenómeno natural que se manifiesta en una zona geográfica específica y de interés para el usuario.

6.2.1.2 Descripción de la Sistema

Se estableció que el SIG debía ser publicado en Web y construido a partir de *software libre*. La aplicación se compone de una interfaz gráfica de usuario, mediante la cual fuese posible realizar un análisis geoestadístico -según unas etapas claramente definidas- de forma automática, construyendo funciones de procedimientos con algoritmos en geoestadística, que operan en una base de datos espacial; los resultados obtenidos ejecutando dichas funciones son alojados en un servidor. La aplicación Web está basada en distintas técnicas de análisis espacial, usadas en las ciencias aplicadas, para la generación de estadísticas descriptivas, gráficos estadísticos, semivariogramas (experimentales y teóricos) e información geográfica continua (mapas de predicción y mapa de varianza de error), que describirían un fenómeno natural en particular.

El sistema completo fue dividido en una serie de módulos básicos y específico, dentro de los básicos se consideraron el módulo de acceso de usuario, módulo de registro de usuario, módulo de documentación, módulo de carga de datos, módulo de gestión de capas geográficas, módulo de lectura, inserción, actualización y borrado de datos (*CRUD*¹⁴); en tanto a los específicos cumplen con el desarrollo de cada una de las etapas del análisis geoestadístico (análisis exploratorio de datos, el análisis estructural de datos y predicción espacial) estos su vez interactúan entre sí para lograr una funcionalidad final.

6.2.1.3 Alcances del sistema

Dado que el SIG fue diseñado con el propósito de realizar un procesamiento geoestadístico a una variable regionalizada para la estimación de un fenómeno, durante su construcción se omitieron algunas capacidades que poseen otros *software* tipo escritorio. Algunas de las limitaciones y alcances del sistema se mencionan a continuación:

¹⁴ **CRUD**, En computación CRUD es el acrónimo de Crear, Obtener, Actualizar y Borrar (del original en inglés: Create, Read, Update and Delete).

Aunque el análisis de anisotropía fue tenido en cuenta, este solo fue realizado para direcciones predeterminadas (45, 90, 135 y 180 grados) y con una tolerancia fija.

No fueron tenidas en cuenta etapas como el análisis bivariado y multivariado.

Aunque existen gran cantidad de semivariogramas teóricos, durante la construcción del SIG únicamente se seleccionaron los más conocidos y utilizados (esférico, exponencial, gaussiano, lineal y efecto pepita puro).

El ajuste del semivariograma experimental a un modelo teórico lo puede realizar el aplicativo de manera automática o se puede dejar a criterio de usuario (según un conocimiento a priori) seleccionando cual de dichos modelos representa mejor el comportamiento de la variable.

Los resultados obtenidos al finalizar la etapa de predicción se presentan en formato raster, por motivos de eficiencia en los cálculos y control de fallos del aplicativo, el usuario no puede modificar el tamaño de pixel en el raster, dicho tamaño lo calcula el sistema realizando una operación o ecuación que relaciona el área de estudio y número de puntos.

A continuación se definen los materiales y procedimientos considerados en este proyecto, se describen también los recursos hardware y software utilizados.

6.2.1.4 Análisis de Requerimientos

Se definieron los requerimientos funcionales del sistema de información geográfico con el fin de establecer las restricciones y comportamientos del aplicativo.

Requerimientos no Funcionales del Sistema

Se especificaron las propiedades del sistema, tales como restricciones del ambiente y desarrollo, performance, dependencias de plataformas integridad y seguridad, estos requerimientos se describen en la tabla 4.

Tabla 4. Listado de Requerimientos no funcionales del sistema

REQUERIMIENTOS NO FUNCIONALES		
Id	Nombre	Descripción
1	Naturaleza	Debe ser un aplicativo Web.
2	Concurrencia	Debe permitir el acceso concurrente.
3	Cliente/Servidor	Debe seguir el modelo arquitectónico cliente/servidor, centralizando los datos consultados en un único lugar (sin replicación).
4	Open Source	Debe ser desarrollada mediante aplicaciones gratuitas, de código libre.
5	Integridad	Debe ofrecer garantías en la integridad de los datos almacenados.
6	Seguridad	Debe ser una aplicación segura, respecto a los datos de los usuarios y los datos propios de la aplicación.
7	Flexibilidad	Debe poder adaptarse a nuevos requisitos funcionales (ser modular).
9	Estabilidad	Debe tener un nivel muy bajo de fallos.
10	Plataforma	Debe ser multiplataforma.
11	Eficiencia	Debe ser eficiente en tiempos de computación.
12	Documentación	Debe disponer de una documentación extensa y adecuada, tanto para usuarios.

Requerimientos Funcionales del Sistema

Al mismo tiempo se definieron una serie de requerimientos funcionales, que establecen los lineamientos de operatividad que el sistema de Información Geográfico debía ofrecer a un usuario (Tabla 5).

Tabla 5. Listado de Requerimientos Funcionales del sistema

REQUERIMIENTOS FUNCIONALES		
Id	Nombre	Descripción
1	Datos	Debe recuperar datos alfanuméricos y cartográficos de una base de datos espacial.
2	Representación	Debe representar datos estadísticos y geoestadísticos a partir de la información base.
3	Registro	Debe ofrecer la opción de registrar usuarios.
4	Identificación	Debe identificar usuarios.
5	Consulta Geoestadística	Debe realizar consultas a variables ambientales
6	Fuentes	El usuario puede ingresar datos y a la vez leer, actualizar y borrar las fuentes de los datos almacenados.
7	Log-out	El usuario debe poder cerrar sesión en cualquier momento.

8	Interfaz	La interfaz Web debe ser sencilla e intuitiva, independiente del diseño.
9	Geo-procesos	Las respuestas deben ser presentadas en gráficos, formato raster y valores cuantitativos en forma de tabla.
10	Visor Cartográfico	Debe de disponer de controles que permitan la interacción con la información espacial.

6.2.1.5 Descripción de Hardware

En el desarrollo del proyecto se trabajó en un computador de escritorio, en este se realizó la implementación y prueba del aplicativo web. Las características técnicas del equipo se muestran a continuación:

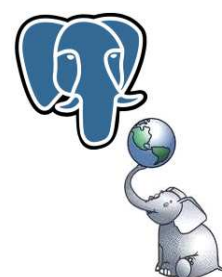
Ficha Técnica	
Procesador	Core2 Duo
Disco Duro	40 Gb
Memoria	3,25 de Ram
Velocidad	2,93 Ghz
Sistema Operativo	Windows 7

6.2.1.6 Selección de Herramientas

Las herramientas en *software* consideradas, funcionan de acuerdo a una arquitectura cliente-servidor; se tuvo en cuenta diversas herramientas considerando los niveles de requerimientos previamente establecidos (Tabla 4.), además se seleccionaron como resultado de un proceso de exploración y revisión bibliográfica que identificó funcionalidades, estándares, eficiencia, fuentes de documentación y facilidad de integración entre las diferentes alternativas, que permitieran desarrollar la metodología propuesta, resultando las siguientes:

- **Servidor de Base de Datos Geográficos:**
PostgreSQL¹⁵ + Postgis¹⁶

El sistema gestor de base de datos (SGBD) elegido para este proyecto fue PostgreSQL + Postgis, el cual gestiona y manipula la información almacenada de forma objeto relacional a partir de PgAdmin (cliente que permite la administración de la información), Postgis actúa como el módulo espacial para el SGBD. Este proporciona a PostgreSQL la capacidad no sólo de almacenar información geoespacial, sino de realizar operaciones de análisis a datos geográficos.



¹⁵ PostgreSQL www.postgresql.org/

¹⁶ PostGIS, módulo soporte datos espaciales para PostgreSQL; <http://postgis.refractory.net/>

- **Servidor de Geoprocesos: R - project¹⁷**

R se utilizó como la herramienta para el procesamiento de la información, este actúa como un sistema para análisis estadísticos generación de gráficos con base a datos espaciales. R tiene una naturaleza doble de programa y lenguaje de programación y es considerado como un dialecto del *lenguaje S* creado por los Laboratorios AT&T Bell. Para el uso de este *software* se provee al gestor de base de datos (PostgreSQL) un lenguaje de procedimiento adicional *pl/R*¹⁸ el cual provee la capacidad de soportar funciones en lenguaje R.



- **Servidor: MS4W¹⁹**

Se eligió MS4W ya que este se orienta a la creación de aplicaciones SIG mediante la tecnología Internet Map Server (*IMS*²⁰). Actúa como paquete pre-configurado en un *sistema operativo Window*, que instala componentes como: el servidor Apache, *Mapserver*²¹ CGI, *PHP*²² y otras utilidades. PHP fue empleado como lenguaje intérprete en el servidor este recuperó y envió la información de la base de datos al cliente; para la ejecución y representación de dicha información de forma asincrónica se utilizó *AJAX*²³.



- **Cliente Web: Mapfish²⁴**

Se optó por Mapfish como framework javascript ya que se caracteriza por ser muy flexible y completo para el desarrollo de aplicaciones de mapeo Web muy amigable para un usuario final además de ser compatible con Ruby on Rails y PHP. Dicho Framework se compone de los siguientes proyectos: OpenLayers, GeoExt, ExtJs además de características propias.



¹⁷ The R Project for Statistical Computing <http://www.r-project.org/>

¹⁸ PL/R lenguaje de procedimiento para PostgreSQL <http://www.joeconway.com/plr/>

¹⁹ MS4W <http://mapserver.org/>

²⁰ IMS, sigla referida a Internet Map Service, traducida al español como Servidor de mapa en Internet.

²¹ MapServer <http://mapserver.org/>

²² PHP lenguaje de programación del lado servidor <http://www.php.net/>

²³ Ajax acrónimo de Asynchronous JavaScript And XML (JavaScript asíncrono y XML), es una técnica de desarrollo web para crear aplicaciones interactivas o RIA (Rich Internet Applications).

²⁴ Mapfish <http://www.mapfish.org/>

6.2.2 Fase de Diseño

Debido a que el SIG fue diseñado con finalidades específicas, como la captura, manipulación, procesamiento y análisis geoestadístico de información y la obtención de superficies de predicción sobre la caracterización de variabilidad espacial del fenómeno analizado, se diseñó un subsistema dividido en módulos funcionales. El sistema fue construido siguiendo las etapas principales del análisis geoestadístico, dichas etapas se describen a continuación:

6.2.2.1 Análisis Exploratorio de Datos

El análisis exploratorio de datos (AED), se fundamenta en el análisis estadístico sobre datos espaciales, se consideraron los siguientes cálculos: valor máximo, mínimo, media, mediana, asimetría, curtosis, desviación estándar y coeficiente de variación para cada una de las variables consideradas. Además de utilizar AED clásico, al disponerse de información georreferenciada se construyeron gráficos como el histograma, el diagrama de caja, el gráfico Q-Q normalizado, el gráfico de dispersión, el diagrama de nubes de semivariograma que permiten conocer la centralidad y dispersión, distribución de los datos e identificar las tendencias en el valor promedio de una variable en la región (relación entre variable medida y coordenadas geográficas), para efectos descriptivos.

6.2.2.2 Análisis Estructural de Datos

Mediante el análisis estructural de datos es posible generar semivariogramas experimentales que determinan la estructura de autocorrelación espacial de la variable analizada, el conocimiento de este permite inferir en la formulación de la predicción espacial. Puesto que existen procesos para los cuales la variación de valores no se da igual en todas las direcciones, se ofrece la opción de generar semivariogramas anisotrópicos que no solo indican la variación media dentro de un intervalo de distancia, sino que caracterizan esa variación para una distancia y una dirección concreta.

A partir de los puntos que forman el semivariograma experimental (definido por la función de *semivarianza*), puede definirse un modelo que aporta información sobre el proceso subyacente, a partir de su forma y sus parámetros²⁵. La definición de este modelo implica el ajuste de una curva a los puntos del semivariograma experimental, y tiene como resultado la obtención de un semivariograma teórico. En la figura 7., puede verse junto a los puntos del semivariograma experimental una curva ajustada a estos que define el semivariograma teórico.

²⁵ Véase, apartado marco conceptual parámetros del semivariograma (*Método Kriging*).

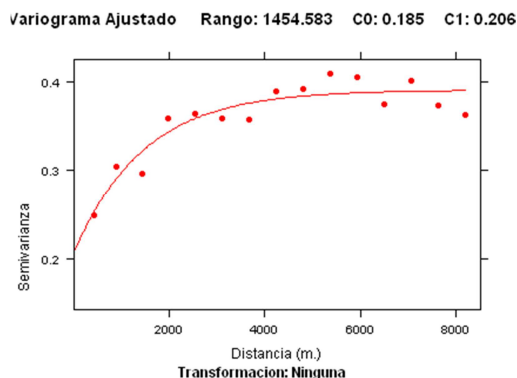


Figura 7. Ajuste del semivariograma teórico.

Fuente: propia.

El aplicativo al soportar modelos matemáticos que describen un semivariograma teórico permite calcular semivarianzas para cualquier distancia h , no solo para las definidas por los intervalos como en el caso del semivariograma experimental.

Las alternativas para elegir funciones que describan el semivariograma teórico van sujetas a los siguientes requisitos:

- Ser monótona creciente.
- Tener un máximo constante o asintótico. Es decir, un valor definido del sill.
- Funciones no acotadas superiormente, tales como las exponenciales, indicarán que la zona de estudio no es suficientemente grande, ya que no alcanza la dimensión a partir de la cual el efecto de la dependencia espacial deja de existir.

Sobre la base anterior, se establecieron las familias principales de funciones aptas para definir un semivariograma teórico:

- Nugget puro. Funciones constantes de la forma

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ C_0 & h > 0 \end{cases}, \text{ donde } C_0 > 0$$

Ecuación 8. Función de modelo Nugget puro.

En este caso, la aplicación del Kriging no resulta posible, y el valor estimado en un punto es la media de los valores muestreados.

- Funciones que alcanzan el valor del sill (c) para un rango concreto(a). Son funciones de la forma.

$$\gamma(h) = \begin{cases} f(x) & \text{si } h \leq a \\ c & \text{si } h > a \end{cases}$$

Ecuación 9. Función de modelo lineal, circular o esférico.

Siendo las más habituales de las funciones $f(x)$ las de tipo lineal, circular o esférico.

- Funciones que tienden asintóticamente al valor del sill (c). Se define un rango efectivo, siendo este en cual la semivarianza es igual a $0.95c$. Entre estas funciones destacan:

a.) $\gamma(h) = c \left(1 - e^{-\frac{3h}{a}} \right)$

b.) $\gamma(h) = c \left(1 - e^{-\frac{9h^2}{a^2}} \right)$

Ecuación 10. Función de modelos a.) Modelo Exponencial; b.) Modelo Gaussiano.

El empleo de uno u otro modelo dependerá del conocimiento que tenga el usuario acerca del proceso modelizado. La figura 8, muestra las gráficas de los tipos de semivariograma anteriores. Llevar a cabo el ajuste del semivariograma teórico no es en absoluto un proceso trivial. Lo más sencillo es tratar de minimizar el error cuadrático.

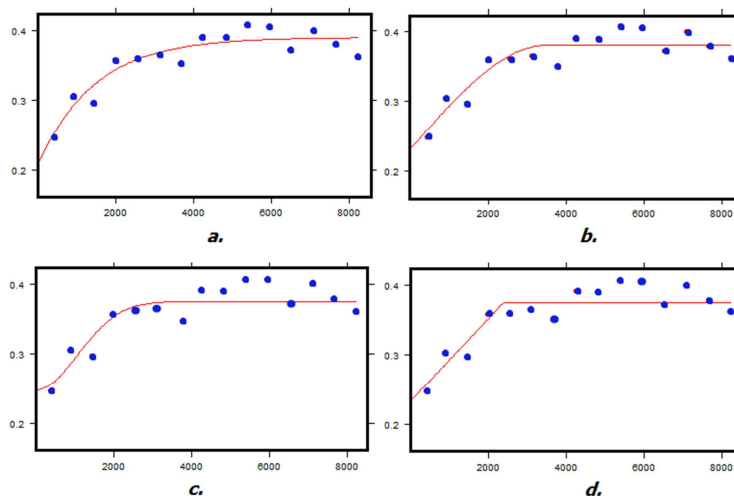


Figura 8. Semivariogramas teóricos soportados por el aplicativo web. a. Modelo Exponencial; b. Modelo Esférico. Modelo Gaussiano; d. Modelo Lineal.

Fuente: propia.

6.2.2.3 Predicción Espacial

Para esta etapa se consideró la generación de información tipo raster con una estructura regular, según sus características es la que mejor representa las variables que tiene un comportamiento continuo en el espacio, en este caso se

aplicó a las propiedades químicas de suelo tomadas en la zona de estudio, ya que el cálculo de dicha información es basada en información irregularmente distribuida en el espacio, se concibió la aplicación de métodos especiales como lo son la interpolación.

El método de interpolación aplicable por el SIG con la finalidad fundamental de análisis espacial de variables ambientales debe ser complejo con fuerte carga geoestadística, fue por esto que el aplicativo Web, fue diseñado con la capacidad de predicción espacial a partir del interpolador Kriging definiendo cada uno de sus tipos de variantes. Dicho método ofrece una forma objetiva de establecer la ponderación óptima entre los puntos en un interpolador local²⁶.

La interpolación óptima aplicada por el SIG, cumple los siguientes requisitos:

- El error de predicción debe ser mínimo.
- Los puntos cercanos deben tener pesos mayores que los lejanos.
- La presencia de un punto cercano en una dirección dada debe restar influencia (enmascarar) a puntos en la misma dirección pero más lejanos.
- Puntos muy cercanos con valores muy similares deben «agruparse», de tal forma que no aparezca sesgo por sobre-muestreo.
- La estimación del error debe hacerse en función de la estructura de los puntos, no de los valores.

El aplicativo SIG fue diseñado para que junto con la superficie interpolada (mapa de predicción), se genere así mismo superficies con medidas del error de interpolación a partir del mapa de varianza, con el propósito de definir la bondad de la predicción en toda la zona estudiada.

6.2.3 Fase de Implementación

A continuación se presentan algunos aspectos tenidos en cuenta durante la implementación del SIG, se presentan también los resultados obtenidos durante la construcción de la base de datos y la interfaz gráfica.

6.2.3.1 Estructura de funcionamiento

El SIG reside en un servidor (*Apache*) al cual se accede mediante el uso de un navegador de Internet (*Mozilla Firefox, Opera, Internet Explorer, etc.*). La interacción de usuario se realiza mediante páginas web diseñadas para trasladar peticiones POST y GET del usuario al servidor que interacciona con la base de

²⁶Véase, apartado marco conceptual (*Interpolación*).

datos *PostgreSQL+Postgis* mediante ODBC²⁷. El usuario común accede al SIG teniendo la facultad de cargar, consultar, procesar, descargar, o editar sus datos, un usuario administrador tendrá adicionalmente la capacidad de actualizar borrar o insertar cualquier función, tabla o esquema de la base de datos espacial.

Los datos a analizarse por medio del aplicativo Web se cargan al aplicativo por medio de archivos planos en formatos con extensión *txt* o *csv*, siempre que estos cumplan con las siguientes condiciones:

- El encabezado en el archivo es obligatorio; dicho encabezado corresponde a los nombres de las columnas de datos que conforman dicho archivo.
- Debe existir obligatoriamente una columna de nombre “x” y otra de nombre “y” que corresponden respectivamente a las coordenadas Este y Norte de los datos, las demás columnas pueden poseer cualquier nombre.
- Los nombres de las columnas no deben presentar comillas simples (‘’) ni dobles (’’’’).
- El separador de campos debe ser la coma (,).
- El separador de decimales debe ser el punto (.).
- No deben existir campos de tipo texto, todos deben ser de tipo numérico.
- No deben existir filas de datos repetidas en el archivo.
- Para cargar los datos al aplicativo es necesario conocer el código EPSG perteneciente al sistema de referencia asociado a los datos.

En la Figura 9 se observa una imagen de un archivo en formato *csv*, en el cual se cumplen las condiciones establecidas anteriormente.

²⁷ **Open DataBaseConnectivity (ODBC)** es un estándar de acceso a bases de datos desarrollado por [SQL Access Group](#).

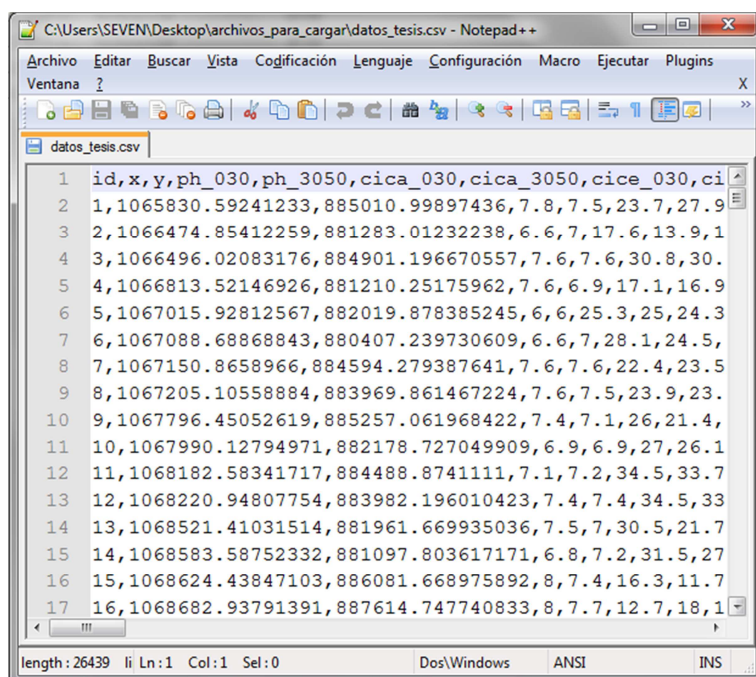


Figura 9. Archivo muestra en formato csv.

En caso de no poseer información en los formatos mencionados, es posible utilizar a manera de prueba los datos que almacena el aplicativo por defecto, cabe decir que según los requerimientos no funcionales del sistema se ofrece garantías en la integridad de los datos almacenados. El acceso a los datos va restringido dependiendo al usuario ya que la base de datos controla el registro de usuarios, lo cual se realiza mediante el uso de identificadores de usuario y contraseñas, y la gestión de sesiones de trabajo por parte del servidor.

De acuerdo a la descripción realizada en las secciones anteriores el esquema global del aplicativo SIG Web, se puede describir en la figura 10.

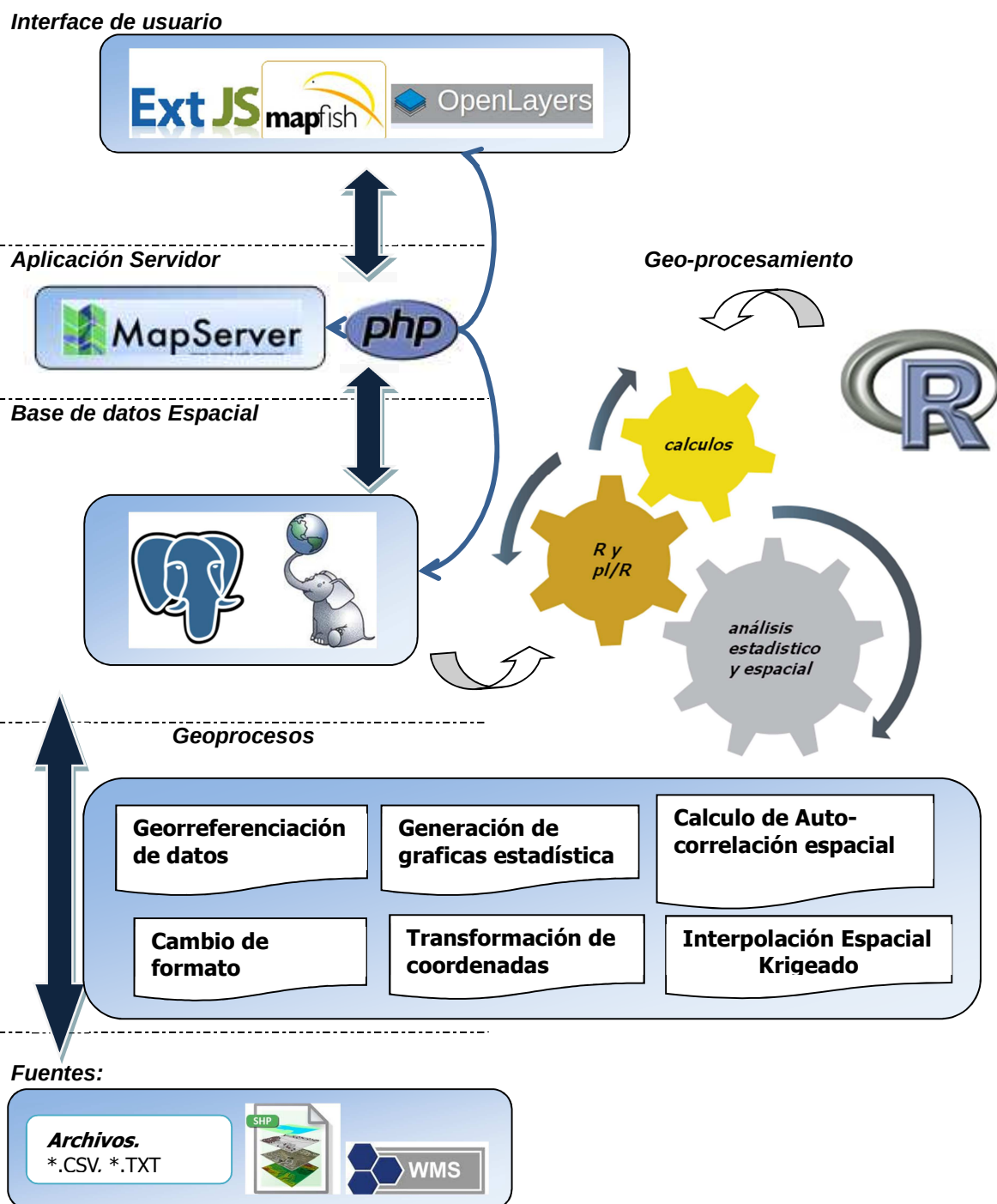


Figura 10. Esquema global del sistema. Arquitectura cliente – servidor



Figura 11. Arquitectura cliente - servidor.

Las figuras 10 y 11, ilustran a manera general y concreta la arquitectura “cliente – servidor” del SIG; en donde, el cliente tiene tareas como: recepción validación de datos y despliegue de resultados a través de la interfaz de usuario; mientras el servidor: recepción y procesamiento de las solicitudes del cliente y envío de respuestas al cliente.

6.2.3.2 Construcción de la base de datos

La base de datos espacial se basó en un modelo entidad-relación y fue construida para cumplir las siguientes tareas: manipulación de los datos sujetos a geoprocésamiento, almacenamiento de capas geográficas (ver Tabla 6 -capas vector), que se colocaron como información base para visualización cartográfica según la zona de estudio, almacenamiento de la información de usuarios (nombres, *username*, *password*, datos, fechas de registro etc.) del aplicativo SIG. Adicionalmente, la base de datos espacial se construyó para que cumpliera con algoritmos en geoestadística.

De modo que, fue necesario que el programa con funcionalidades en geoestadística (*software R*) fuera provisto de múltiples paquetes de funciones (ver **Anexo C.**) que cumplieran con finalidades específicas de cálculo y gráficos además que su instalación fuera en la misma máquina en donde se instaló el sistema manejador de base de datos ya que las funcionalidades del SIG radican en la interoperabilidad entre estos dos software.

Se crearon funciones de procedimiento nativas y foráneas (pl/pgSQL y pl/R). Con respecto a las funciones pl/R como su nombre lo indica fueron escritas en lenguaje R, pero interpretadas como funciones propias de la base de datos espacial. Tales funciones se ejecutan mediante órdenes del usuario a través del cliente, la estructura de las funciones permitió ejecutar múltiples cálculos para una sola consulta de usuario, respondiendo a las necesidades de geoprocésamiento.

La estructura de una función de procedimiento almacenada en la base de datos espacial se representa a continuación:

```

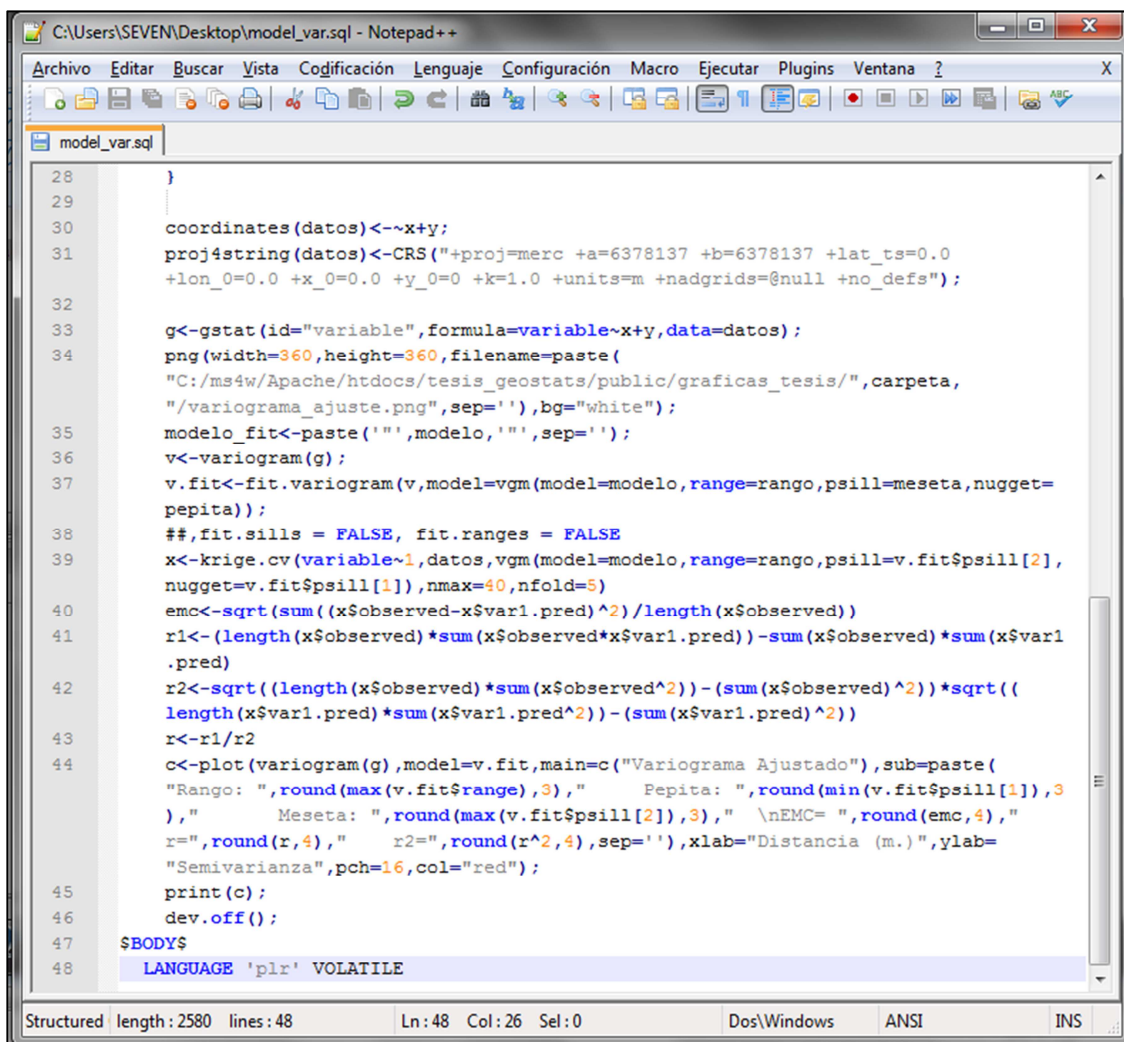
5 CREATE OR REPLACE FUNCTION r_variograma_fit_user(tabla text, variable text, modelo
character varying, meseta character varying, rango character varying, pepita
character varying, trans text, lamb double precision, carpeta character varying)
6 RETURNS text AS
7 $BODY$
8     library(gstat);
9     library(car);
10
11     consulta<-paste("SELECT x(st_transform(the_geom,900913)) as x,
y(st_transform(the_geom,900913)) as y, ",variable," as variable FROM ",tabla,"
WHERE ",variable,"!=0",sep=' ');
12
13     datos<-pg.spi.exec(consulta);
14
15     if (trans=="log"){
16     datos$variable<-log(datos$variable);
17     t<-"Logaritmica";
18     }
19
20     if (trans=="boxcox"){
21     datos$variable<-bcPower(datos$variable,lamb=lamb);
22     t<-paste("Box-Cox"," ",lamb,"Lambda: ",lamb,sep=' ');
23     }
24
25     if (trans=="none"){
26     datos$variable<-datos$variable;
27     t<-"Ninguna";
28     }
29
30     coordinates(datos)<-~x+y;
31     proj4string(datos)<-CRS("+proj=merc +a=6378137 +b=6378137 +lat_ts=0.0
+lon_0=0.0 +x_0=0.0 +y_0=0 +k=1.0 +units=m +nadgrids=@null +no_defs");

```

Figura 12. Función de procedimiento en R. para cálculo y ajuste de un semivariograma Experimental. (parte1).

Basados en la figura 12 y 13, se puede observar que a diferencia de una sentencia o consulta estándar en una base de datos, una función permite realizar múltiples operaciones orientadas a generar uno o varios resultados, en las figuras mencionadas se visualiza la estructura de una función escrita en lenguaje de procedimiento R; en ella se hizo previamente el llamado de paquetes especializados del software R y posteriormente se escribió una rutina para diversos cálculos como selección de coordenadas "x" y "y" en los datos de entrada, aplicación de parámetros de transformación, asignación de proyección (Figura 12.), definición de rutas que definen el lugar de almacenamiento de datos geo procesados, aplicación de modelos matemáticos para ajuste de un variograma, definición de parámetros de ajuste, cálculos de validación (error medio cuadrático *EMC* y coeficiente de determinación r^2), definición de parámetros

a ser impresos, tamaño, estilos, y tipo de formato de salida (Figura 13.).



```

28     }
29
30     coordinates(datos)<-~x+y;
31     proj4string(datos)<-CRS("+proj=merc +a=6378137 +b=6378137 +lat_ts=0.0
+lon_0=0.0 +x_0=0.0 +y_0=0 +k=1.0 +units=m +nadgrids=@null +no_defs");
32
33     g<-gstat(id="variable",formula=variable~x+y,data=datos);
34     png(width=360,height=360,filename=paste(
"C:/ms4w/Apache/htdocs/tesis_geostats/public/graficas_tesis/",carpeta,
"/variograma_ajuste.png",sep=' '),bg="white");
35     modelo_fit<-paste('','',modelo,'',sep=' ');
36     v<-variogram(g);
37     v.fit<-fit.variogram(v,model=vgm(model=modelo,range=rango,psill=meseta,nugget=
pepita));
38     ##,fit.sills = FALSE, fit.ranges = FALSE
39     x<-krige.cv(variable~1,datos,vgm(model=modelo,range=rango,psill=v.fit$psill[2],
nugget=v.fit$psill[1]),nmax=40,nfold=5)
40     emc<-sqrt(sum((x$observed-x$var1.pred)^2)/length(x$observed))
41     r1<-(length(x$observed)*sum(x$observed*x$var1.pred))-sum(x$observed)*sum(x$var1
.pred)
42     r2<-sqrt((length(x$observed)*sum(x$observed^2))-(sum(x$observed)^2))*sqrt((
length(x$var1.pred)*sum(x$var1.pred^2))-(sum(x$var1.pred)^2))
43     r<-r1/r2
44     c<-plot(variogram(g),model=v.fit,main=c("Variograma Ajustado"),sub=paste(
"Rango: ",round(max(v.fit$range),3),"      Pepita: ",round(min(v.fit$psill[1]),3
),"      Meseta: ",round(max(v.fit$psill[2]),3),"      \nEMC= ",round(emc,4),"
r=",round(r,4),"      r2=",round(r^2,4),sep=' '),xlab="Distancia (m.)",ylab=
"Semivarianza",pch=16,col="red");
45     print(c);
46     dev.off();
47     $BODY$
48     LANGUAGE 'plr' VOLATILE

```

Figura 13. Función de procedimiento en R. para calculo y ajuste de un semivariograma Experimental. (parte2).

6.2.3.3 Construcción de la interfaz grafica

La interfaz fue diseñada con una apariencia entendible y fácil de usar, tan parecida como la de una aplicación de escritorio, en ella los elementos fueron dispuestos de forma simple y ordenada. Por ello, se creó una interfaz gráfica en la que se facilitara al usuario la utilización del SIG. Dicha realización fue posible mediante el empleo de *FrameWork's en JavaScript* y *AJAX*, que permitieron crear una aplicación compleja utilizando diversos componentes gráficos como ventanas, grid (tablas), paneles y layouts que fueron representados dinámicamente. Es por esto

que dicha estructura dio como resultado un sistema modular entendible (Figura 14.). En este sentido, los módulos del SIG en Web corresponden a las etapas principales de un análisis geoestadístico convencional (análisis exploratorio, análisis estructural y predicción), como también el módulo de registro, módulo de acceso, módulo de documentación, módulo de edición de datos de usuario y el módulo de descarga de información resultante de geoprocesamiento.

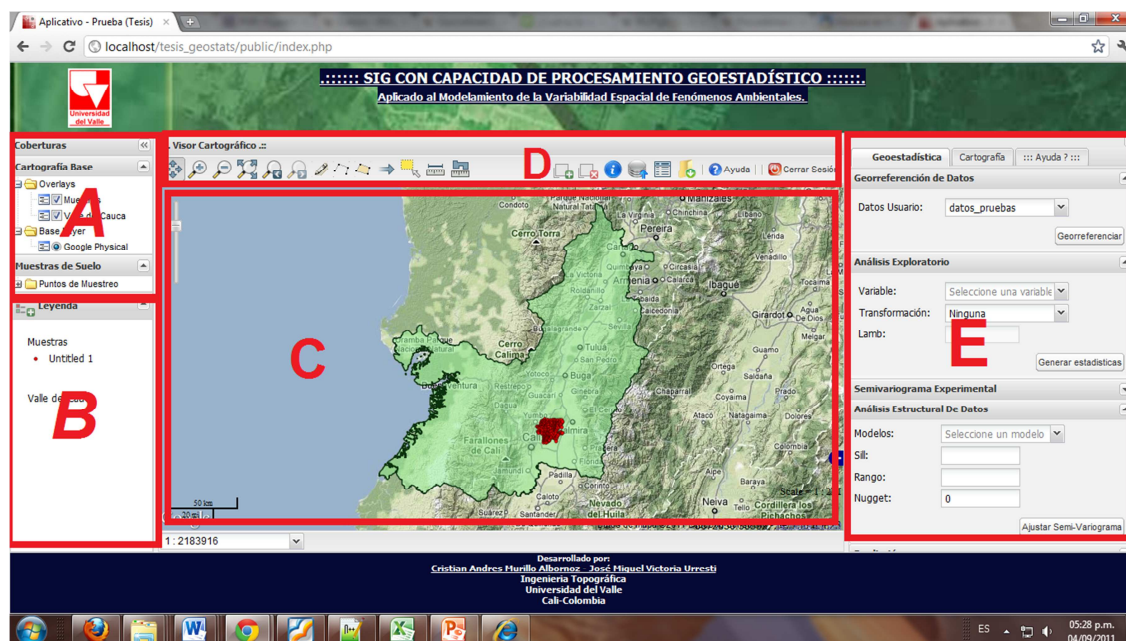


Figura 14. Estructura modular - Interfaz gráfica de usuario.

En la Figura 14 se muestran los componentes principales de la interfaz gráfica, el *recuadro A* corresponde al panel de control de capas, el *recuadro B* representa el panel de leyenda, en la *recuadro C* se encuentra ubicado el visor cartográfico en este se pueden representar diferentes capas geográficas, según la zona de estudio el aplicativo inicialmente utilizo una capas de información base (Tabla 6.) , el *recuadro D* corresponde al panel de controles de usuarios además el módulo de gestión de información espacial, finalmente, en la *recuadro E* se encuentran agrupados los módulos de análisis exploratorio, análisis estructural y predicción espacial.

Tabla 6. Información espacial utilizada.

Nombre	Estructura	Primitiva	S. de Coordenadas
Rio Cauca/Valle del Cauca	vector	Linea	Planas EPSG:21891
Municipios/ Valle del Cauca	vector	Polígono	Planas EPSG:21892
Centros Poblados /Valle del Cauca	vector	Polígono	Planas EPSG:21893
Cuenca Amaime / Valle del Cauca	vector	Polígono	Planas EPSG:21894

Complejo Hidrológico/ Cuenca Amaime	vector	Linea	Planas EPSG:21895
MDE/Valle del Cauca	raster	pixel	Geograficas EPSG:4326
MDE/Mundo	raster	pixel	Geograficas EPSG:4326

Una descripción más detallada de los módulos y paneles se presenta a continuación:

Módulo de acceso: el SIG guarda el registro de usuarios y los datos cargados por estos, de esta manera un usuario únicamente puede manipular sus propios datos, este módulo fue construido con el propósito de controlar el ingreso al SIG a dichos usuarios y determinar los privilegios que cada uno de estos tiene.

El módulo consiste en una ventana mediante la cual se debe ingresar un nombre de usuario y una contraseña, únicamente si dichos datos son correctos es posible acceder a las capacidades del SIG en caso contrario es necesario que el usuario cree un nombre de usuario y contraseña mediante el módulo de registro.



Figura 15. Módulo de acceso a SIG Web.

Módulo de registro: el SIG fue construido con el propósito de garantizar el acceso público, sin embargo, fue necesario crear un registro de usuarios que garantizara la privacidad de los datos ingresados por cada uno de dichos usuarios. Este módulo tiene como propósito el crear un nombre de usuario y contraseña válidos para quien no se encuentre en el registro de usuarios de manera que sea posible acceder al sistema.

Consiste en un formulario en el cual se deben ingresar un conjunto de datos básicos necesarios para la creación del nombre de usuario y contraseña para su posterior almacenamiento en la base de datos.

Figura 16. Módulo de registro a SIG Web.

Módulo de gestión de información espacial: contiene las herramientas básicas de manejo y visualización de información espacial que todo SIG posee, algunas de estas son:

Panel cartográfico mediante el cual se renderiza la información espacial, y despliega una barra de herramientas que controla la interacción con el visor cartográfico; esta posee controles tales como navegación, zoom hacia adentro, zoom hacia afuera, zoom completo, zoom previo, zoom siguiente, medición de distancias, medición de áreas.

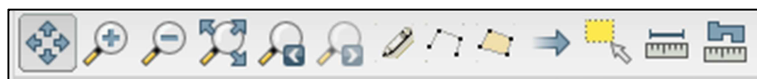


Figura 17. Herramientas del panel visor cartográfico.

El **panel de gestión de datos** mediante el cual es posible acceder a herramientas como: carga o borrado capas geográficas al visor cartográfico, consulta de información alfanumérica mediante un clic sobre el visor cartográfico, carga datos de muestras de usuario, lectura inserción actualización o borrado de datos de usuario, descarga de los resultados de procesamiento, acceso a módulo de ayuda y cierre de sesión de usuario.



Figura 18. Panel de control de usuarios.

El **panel de control de capas** cuya finalidad es la de controlar el orden en el que la información espacial es desplegada a los usuarios, permite también apagar las capas que no se desee visualizar.



Figura 19. Panel de control de capas.

El **panel de leyenda** mediante la cual es posible visualizar el estilo asociado a las capas desplegadas en el panel mapa.

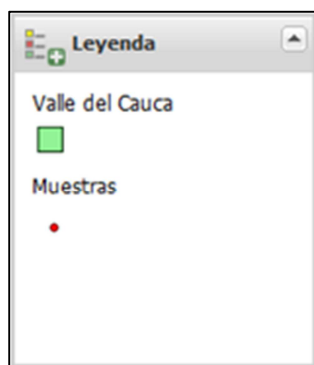


Figura 20. Panel leyenda.

Módulo de análisis exploratorio: permite la selección de la variable a ser analizada en caso de ser conveniente también es posible la aplicación de transformación a los datos. El propósito de este módulo es el de generar las estadísticas básicas y gráficos descriptivos para la variable seleccionada.

Figura 21. Módulo de análisis exploratorio.

Módulo de análisis estructural: permite generar el semivariograma experimental de la variable seleccionada previamente, permite también ajustar dicho semivariograma a un modelo teórico, para esto es posible seleccionar libremente el modelo y los parámetros de ajuste.

Los modelos de ajuste del semivariograma que considera el sistema son: esférico, exponencial, gaussiano, lineal y efecto pepita puro, los parámetros de ajuste mencionados anteriormente son respectivamente el efecto pepita, la meseta y el rango. Cuando se realiza la modelización del semivariograma teórico, el usuario puede proponer diversos modelos que se ajusten a los datos muestrales.

Figura 22. Módulo de análisis estructural de datos.

Módulo de predicción: una vez seleccionada la variable, la transformación de datos (en caso de ser necesaria), el semivariograma teórico y los parámetros de ajuste, este módulo permite realizar el proceso de predicción, para esto, permite seleccionar el tipo de Kriging Lineal (ordinario, simple o universal) que se desea emplear. Como resultado se obtiene un mapa de predicción en formato raster en el que se describe el comportamiento de la variable analizada, un mapa de varianzas en la predicción y el error medio cuadrático (EMC) asociado a la predicción realizada.

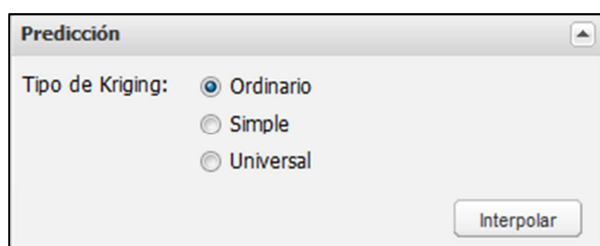


Figura 23. Módulo de predicción espacial – Kriging Lineales.

Módulo de carga de datos. Este módulo permite la carga de datos de usuario a partir de un archivo con formato *txt* o *csv*, se implementó este módulo para realizar la recepción y validación de los datos y de esta forma asegurar el correcto almacenamiento de la información al sistema.

Módulo de descarga. Este módulo permite realizar la descarga de los datos resultado del geoprocesamiento; los resultados se almacenan y agrupan según la etapa geoestadística utilizada. La información que genera el aplicativo comprende archivos en formato *png* que representan de graficas estadísticas generadas en la etapas de análisis exploratorio datos y análisis estructural de los datos, archivos *txt* que contiene información alfanumérica sobre los cálculos geoestadísticos, archivos *shapefile*, *kml* y *tif* que describen información geográfica tipo vector y raster dicha información describe la red de muestreo y las superficies de predicción y de varianza de error respectivamente.

Módulo de documentación: Es un módulo que permite al usuario conocer sobre el proyecto, además se describe en el de forma clara y detallada los procedimientos necesarios para realizar un correcto uso de las funcionalidades del aplicativo SIG - Web. Adicionalmente dicha documentación está dispuesta a ser descargada en formato *pdf* (ver **Anexo G.**).

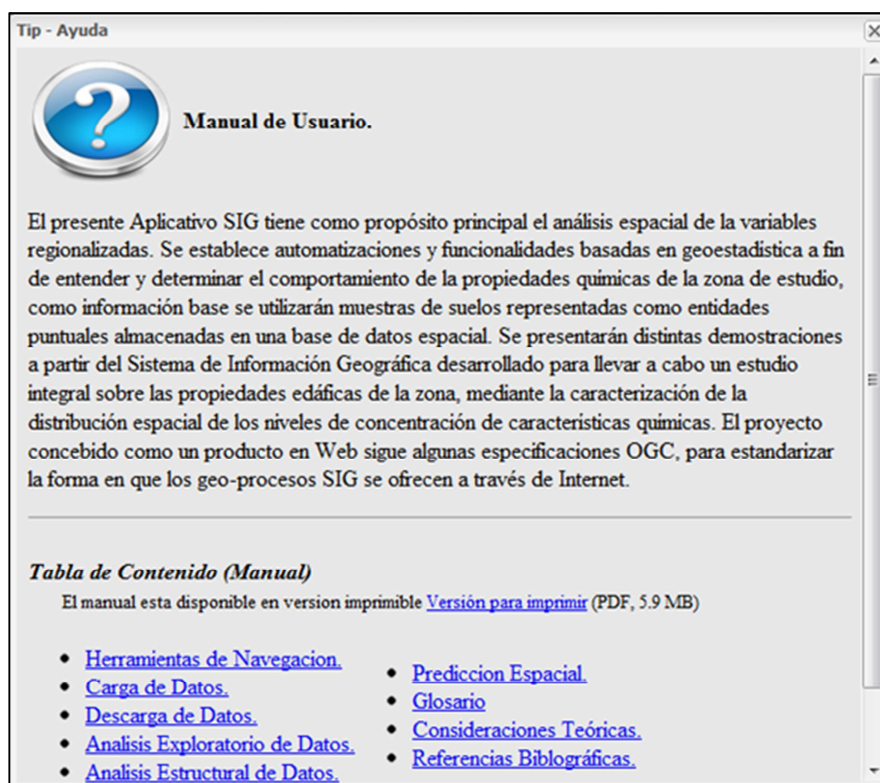


Figura 24. Módulo de ayuda.

6.3 ANÁLISIS GEOESTADÍSTICO

El Aplicativo SIG implementado fue utilizado para realizar el análisis geoestadístico a las variables mencionadas en la sección 6.1.3. Dicho análisis fue realizado de acuerdo a las etapas consideradas en el aplicativo; análisis exploratorio de datos, análisis estructural y predicción espacial.

6.3.1 Análisis exploratorio de datos

Durante el análisis exploratorio se seleccionaron, prepararon y corrigieron los datos antes de ser ingresados al SIG. Una vez ingresados dichos datos, se verificó que estos cumplieran con los supuestos de normalidad y homogeneidad mediante la revisión de las respectivas estadísticas descriptivas y gráficos estadísticos, todos estos generados por el SIG.

Los supuestos de normalidad y homogeneidad de los datos se determinaron teniendo en cuenta las siguientes consideraciones:

- Se consideraron como homogéneas todas las variables cuyo coeficiente de variación (C.V) era menor al 30%, las que poseían un C.V entre 30% y 70% se consideraron ligeramente heterogéneas, finalmente, las variables con C.V mayor al 70% se consideraron heterogéneas.
- Se consideró que las variables presentaban normalidad si la diferencia entre la media y la mediana para la variable analizada era menor que 1 y si la magnitud de su coeficiente de asimetría se encontraba entre 0 y 0.5.

Para las variables que no presentaron normalidad, fue necesario aplicar transformaciones a los datos, lo cual se realizó bajo los siguientes criterios:

- Si la magnitud del coeficiente de asimetría se encontraba entre 0.5 y 1 se realizó una transformación de tipo raíz cuadrada (mediante la familia de transformaciones Box-Cox) a los datos.
- Si la magnitud del coeficiente de asimetría es mayor a 1 se realizó una transformación de logarítmica a los datos.

6.3.2 Análisis estructural de datos

Para este análisis se determinó la estructura de dependencia espacial de las variables mediante la construcción del semivariograma experimental y su posterior ajuste a los modelos teóricos con los que contaba el SIG, según la forma y distribución del semivariograma experimental. Finalmente se seleccionó el mejor de los ajustes con base en el menor error medio cuadrático.

Durante el análisis estructural se identificó si las variables presentaban dependencia espacial, también se identificó el grado de dicha dependencia para lo cual se tuvieron en cuenta las siguientes consideraciones:

Las variables fueron consideradas espacialmente dependientes si el valor de la semivarianza no era constante con respecto a la distancia.

La dependencia espacial se consideró fuerte para todas las variables cuya relación pepita/meseta era menor a 0.25, moderada si dicha relación se encontraba entre 0.25 y 0.75, finalmente, se consideró que la dependencia espacial sería débil para variables cuya relación pepita/meseta era superior a 0.75.

6.3.3 Predicción espacial

Se tuvieron en cuenta los elementos del semivariograma ajustado en la etapa anterior (modelo, pepita, meseta y rango) para generar el mapa de predicción y mapas de varianza para las 7 variables químicas a dos profundidades según la información proporcionada por la corporación autónoma de Valle del Cauca (CVC), ya que las variables se tomaron como estacionarias y la media de la variable es constante pero desconocida se utilizó Kriging ordinario.

7. RESULTADOS

A continuación se presentan los resultados obtenidos con respecto a la ejecución de la funcionalidades del sistema de información geográfica Web, como también los resultados según las reflexiones y análisis sobre al procesamiento geoestadístico realizado a las muestras de suelo seleccionadas como caso de estudio.

7.1 FUNCIONAMIENTO DEL APLICATIVO

A continuación se presentan algunas pruebas realizadas con el fin de evaluar el funcionamiento de los módulos que conforman el SIG, tales pruebas aportaron una mayor comprensión al funcionamiento de dicho aplicativo.

7.1.1 Acceso al aplicativo

Al ingresar al sistema fue desplegada la ventana correspondiente al módulo de ingreso, mediante el cual es posible el ingreso a la ventana principal de la interfaz gráfica una vez introducidos un nombre de usuario y una contraseña. Puesto que el propósito principal de estas pruebas fue el de verificar el funcionamiento del SIG, se creó un nuevo usuario mediante el módulo de registro (Figura 25.).

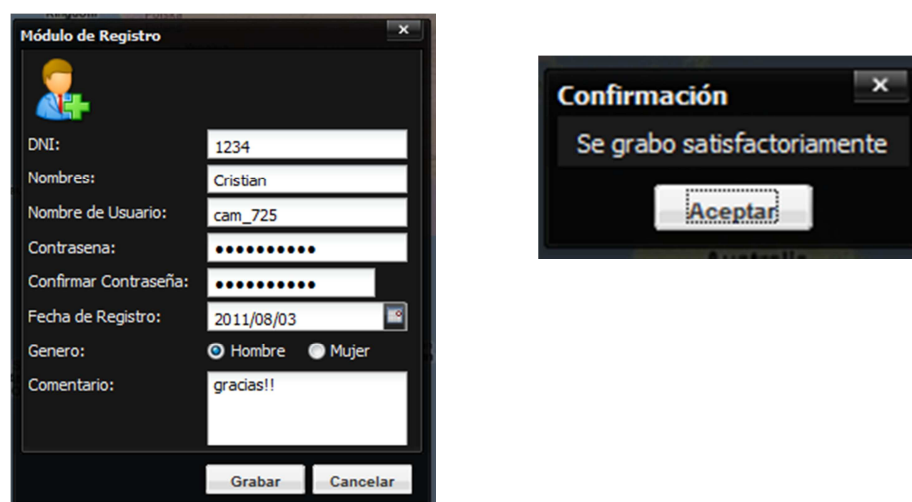


Figura 25. Creación de un nuevo usuario y mensaje de confirmación.

El nombre y contraseña del nuevo usuario fueron utilizados para el ingreso a la ventana principal del sistema. Con el propósito de comprobar el reconocimiento de

usuarios se introdujo en primer lugar una contraseña equivocada, el resultado de fue como era de esperarse un mensaje de error indicando que los datos introducidos no eran válidos (Figura 26).

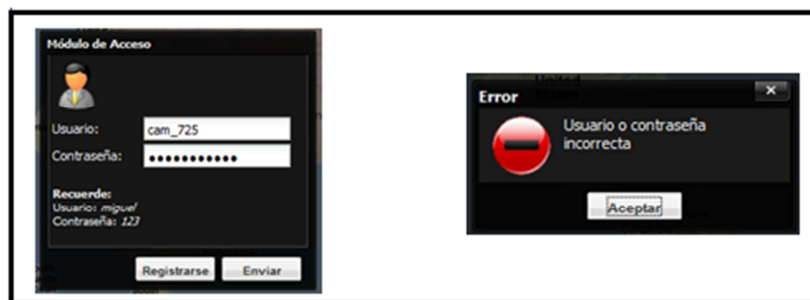


Figura 26. Mensaje de error al ingresar al sistema.

7.1.2 Carga de datos al aplicativo

El módulo para cargar datos fue puesto a prueba, debido a que el sistema está en capacidad de diferenciar los datos de cada usuario, el usuario únicamente puede procesar sus propios conjuntos de datos.

La ventana de ingreso de datos, a la cual se accede mediante el icono () permitió cargar los datos a la base de datos espacial. Para ingresar los satisfactoriamente se debe especificar un nombre, la ruta al archivo que almacena los datos, un código EPSG que define el sistema de referencia y una descripción de los datos; en la Figura 27, se muestran los campos con la información requerida para utilizar la herramienta y el mensaje de aviso indicando que la carga de datos fue exitosa.

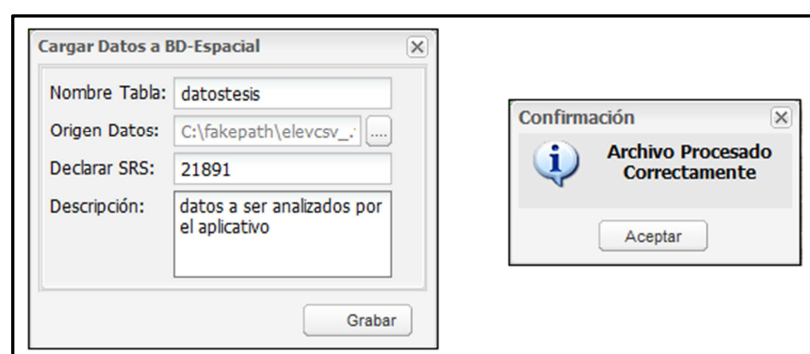


Figura 27. Campos requeridos para cargar un archivo y mensaje de confirmación.

7.1.3 Lectura, actualización y borrado de datos

Con los datos cargados al aplicativo se pudo hacer gestión sobre ellos, el aplicativo permitió seleccionar el conjunto de datos deseado, y de esta manera leerlos y realizar actualizaciones en la información mediante la inserción de un nuevos puntos de muestreo, actualización de valores en específico o el borrado de registro (puntos de muestreo) que no se deseara analizar (Figura 28).

X	y	ph_1	ph_2	calcio_1	calcio_2
1066474.85412259	881283.01232238	6.6	7	10.9	9.4
1066496.02083176	884901.196670557	7.6	7.6	53.3	36.1
1067015.92812567	882019.878385245	6	6	11.8	12.5
1067088.68868	880407.239730609	6.6	7	17.9	12.9
1067150.8658966	884594.279387641	7.6	7.6		29.3
1067205.10558884	883969.861467224	7.6	7.5		22.6

Figura 28. Asistente de edición de datos.

7.1.4 Uso de los módulos de análisis geoestadístico

Se comprobó el funcionamiento de los módulos de análisis geoestadístico mediante el procesamiento de los datos cargados previamente. El paso inicial para dicho procesamiento consistió en la selección de los datos para esta tarea se hizo uso del Menú "Datos de Usuario" (Figura 29), en el cual se encuentra el registro de datos cargados por el usuario, como se mencionó anteriormente por motivos de seguridad e integridad de los datos un usuario únicamente puede visualizar los datos cargados por el mismo, por tal motivo, en dicho menú únicamente serán visibles los datos cargados por el actual usuario.

Figura 29. Menú de selección de datos a analizar.

El botón “Georreferenciar” (Figura 29) fue puesto a prueba para probar la capacidad del SIG de ubicar espacialmente y de manera dinámica el conjunto de datos seleccionado previamente por el usuario (Figura 30).

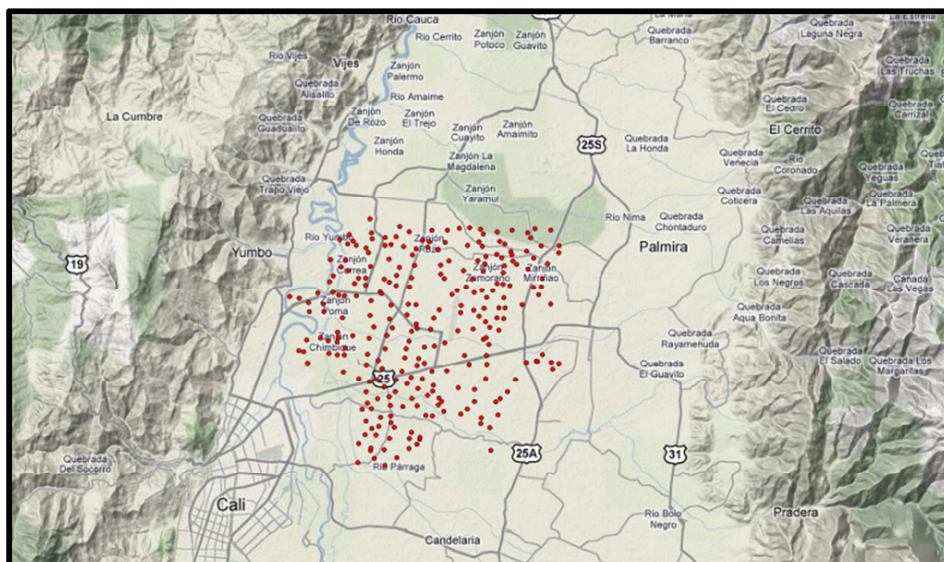


Figura 30. Georreferenciación de datos cargados.

7.1.4.1 Generación de gráficos y de estadísticas descriptivas

Se puso a prueba el módulo de análisis exploratorio mediante la generación de gráficos y estadísticas básicas, para esta tarea fueron utilizados los datos correspondientes a la variable “magnesio_030” perteneciente a los datos cargados previamente. La Figura 31, muestra los parámetros necesarios para llevar a cabo dicho análisis, los cuales corresponden a la variable a analizar (1), la transformación a aplicar a los datos (2) y el parámetro lambda en el caso de las transformaciones Box-Cox (3).

Figura 31. Parámetros necesarios para el análisis gráfico y estadístico.

Las Figuras 32 y 33 muestran los resultados obtenidos al utilizar el módulo en cuestión, se muestran las estadísticas calculadas y los gráficos generados.

Estadísticas Descriptivas								
N_puntos	Mínimo	Maximo	Media	Mediana	Desviación	Asimetría	Kurtosis	C.V.
262	0.2	28.6	10.55	9.75	5.63	0.6925	0.04	53.3649...

Figura 32. Resultados del análisis estadístico.

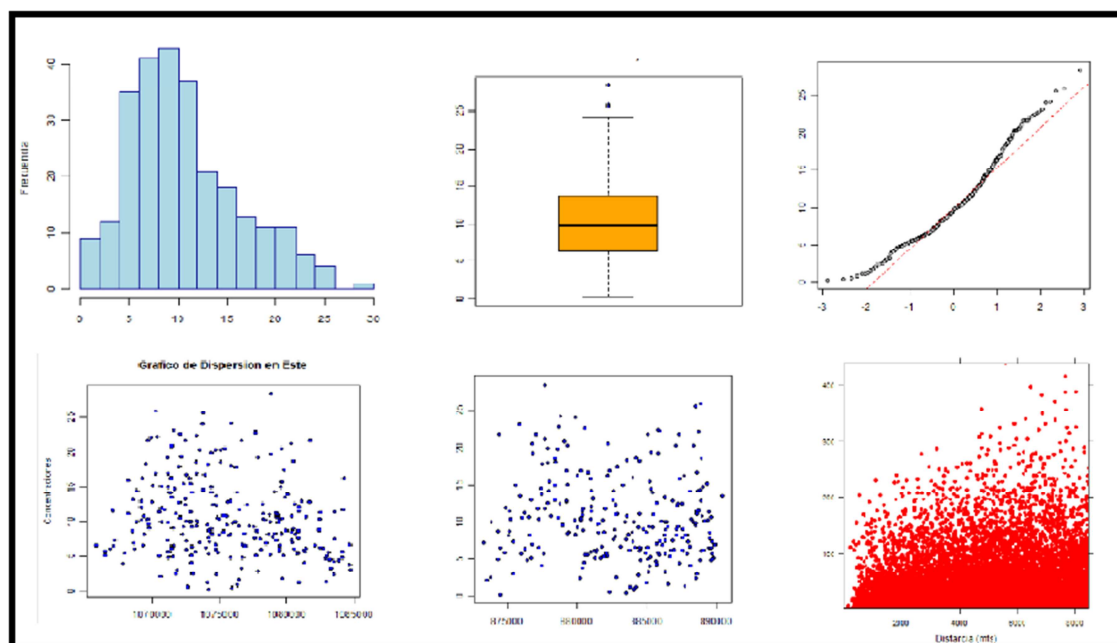


Figura 33. Gráficos estadísticos generados.

7.1.4.2 Generación y ajuste del semivariograma experimental

Para la prueba realizada al módulo de análisis estructural fueron establecidos los parámetros utilizados en la prueba anterior y se generó el semivariograma experimental y se ajustó este a un modelo teórico. Para realizar dicha prueba fue necesario seleccionar los siguientes parámetros: semivariograma sin anisotropía (Figura 34) en el caso del semivariograma experimental. Para el ajuste del semivariograma a un modelo teórico los parámetros fueron: el modelo teórico (a), la meseta (b), el rango (c) y la pepita (d). El resultado genera semivariogramas, como se observa en la Figura 35.

Figura 34. Parámetros necesarios para la generación y ajuste del semivariograma experimental.

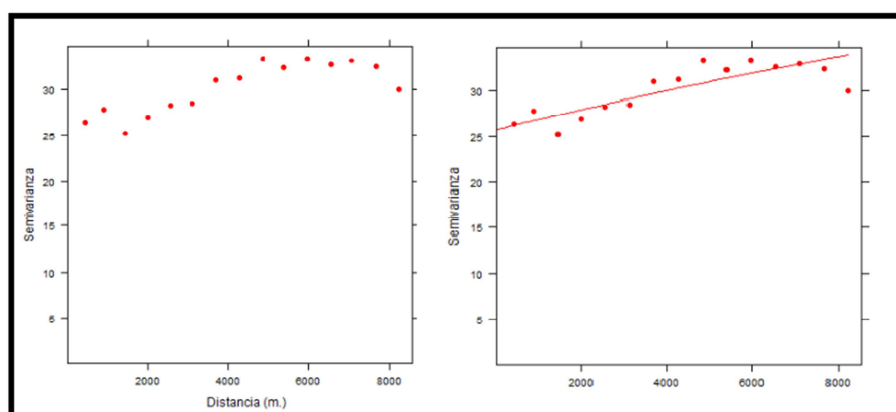


Figura 35. Semivariograma experimental (izquierda) y ajustado (derecha)

7.1.4.3 Funcionamiento módulo de Predicción Espacial.

Para la prueba al módulo de predicción espacial fueron reutilizados los parámetros establecidos en pruebas anteriores (análisis exploratorio y análisis estructural), para poner en funcionamiento dicho módulo fue necesario seleccionar la opción correspondiente al método del interpolador Kriging Lineal con el que se deseaba realizar la interpolación (Figura 36).



Figura 36. Selección del tipo de Kriging.

En la Figura 37, se puede visualizar la capa raster generada después de usar cada una de las etapas geoestadísticas presentes en el aplicativo SIG; el geoprocetamiento realizado corresponde a la variable de acidez (pH) en la zona de estudio.

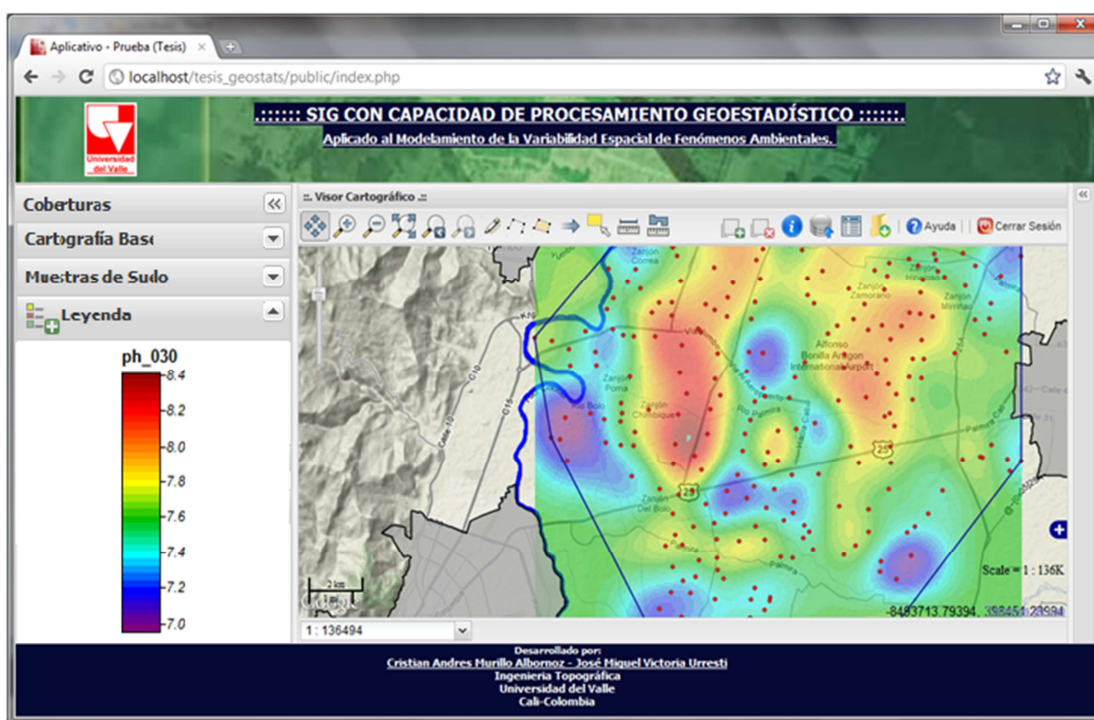


Figura 37. Superficie raster sobre la variable acidez (pH). Vista sobre el visor cartográfico.

Adicionalmente se genera un mapa de varianza que da información acerca de la bondad de la predicción especial sobre la variable que se ha estimado. Se puede observar los errores más altos según la Figura 38, se dan en las zonas en donde la densidad puntos de observación es poca o en zonas por fuera del área de estudio

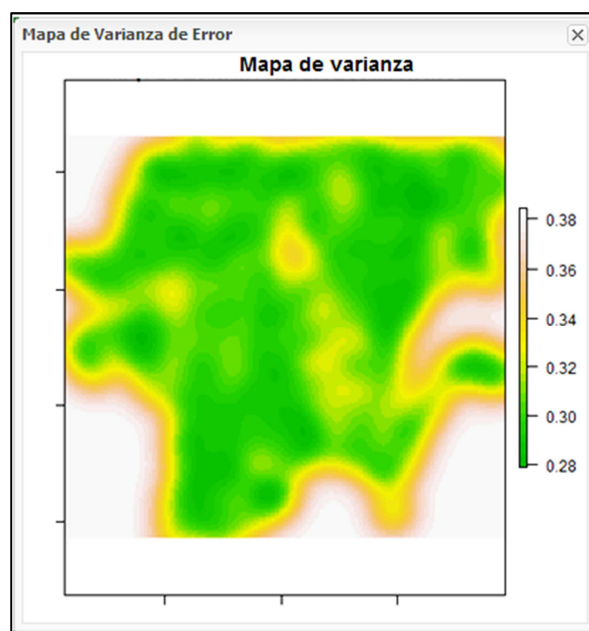


Figura 38. Mapa de varianza de error.

7.1.5 Descarga de información geo-procesada

Una vez utilizado satisfactoriamente cada uno de los módulos de análisis geoestadístico, utilizando como única fuente de información un archivo *csv* o *txt*, que contenía información sobre un levantamiento tipo muestreo, se colocó a prueba el módulo de descarga de datos. Se pudo descargar los resultados del geoprocesamiento; los productos generados abarcan: graficas de estadísticas descriptivas, graficas que definen la estructura correlación espacial de la variable analizada (variogramas experimentales y teóricos), datos tipo raster, como el mapa de predicción y mapa de varianza de error y datos tipo vector como archivos shapefile y archivos *km,l* estos datos describe la red de muestreo.

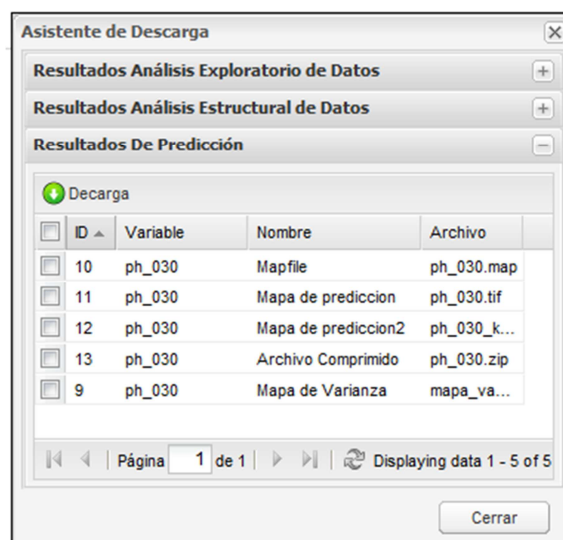


Figura 39. Asistente para descarga de datos.

La información raster o vector estaba debidamente georreferenciada de acuerdo al sistema de referencia ingresado por el usuario como también las coordenadas descritas en el archivo .csv o .txt cargado al sistema. Dicha información pudo ser manejada en cualquier software GIS de escritorio, se puede observar en las figuras 40. y 41, que dicha información se representa con un sistema de coordenadas y datum geográfico de forma correcta, para este caso se visualizó en el software Quantum-GIS²⁸ versión 1.2.0 Tethys.

²⁸ Quantum GIS (o QGIS) es un Sistema de Información Geográfica (SIG) de código libre para plataformas GNU/Linux, Unix, Mac OS y Microsoft Windows. <http://www.qgis.org/>

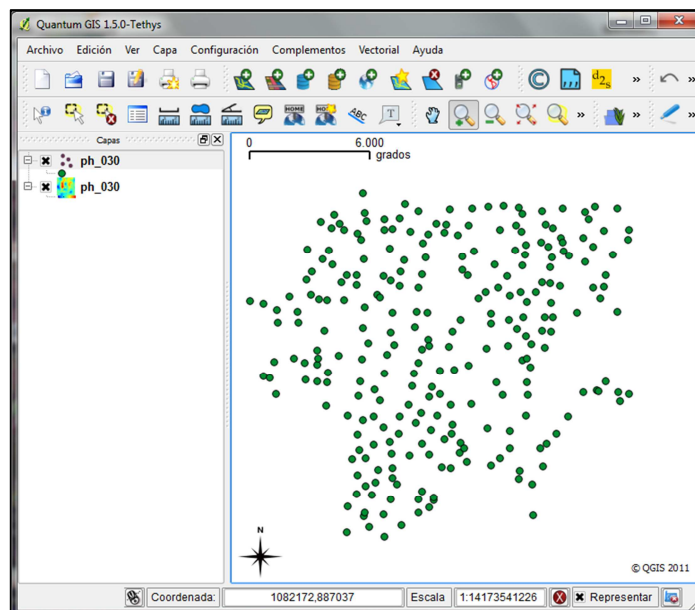


Figura 40. Red de muestreo (propiedades químicas de suelo sector río Fraile río Amaime). Visualización en QGIS.

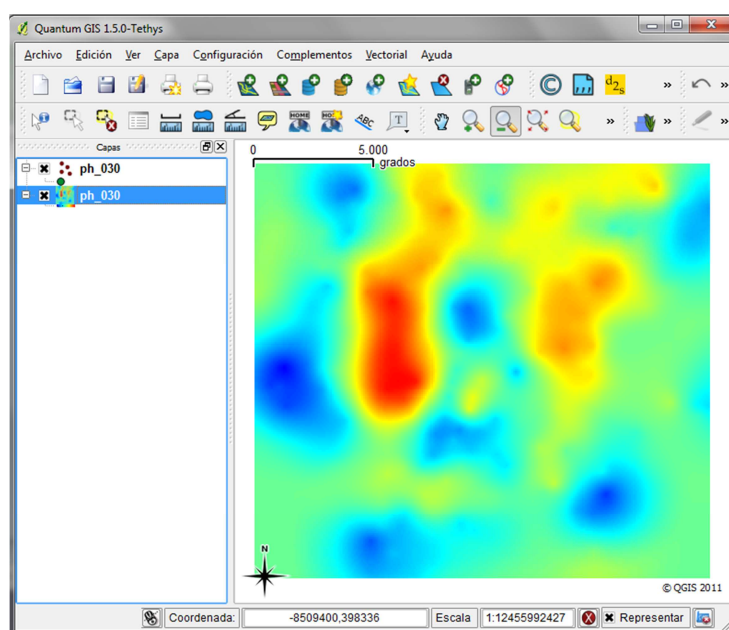


Figura 41. Superficie raster, superficie de predicción sobre la variable pH profundidad 0-30cm.

7.1.6 Validación de Información ingresada por el usuario

Se comprobó el funcionamiento de los mensajes y las advertencias del sistema mediante la omisión de determinados valores al momento de realizar determinadas tareas en el SIG.

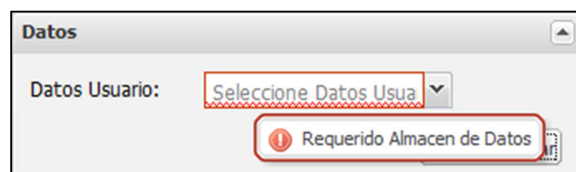


Figura 42. Advertencia generada por el sistema.

La Figura 42, muestra la advertencia generada por el sistema al presionar el botón “Georreferenciar” sin haber indicado antes el conjunto de datos deseado. Advertencias similares se desplegaron al presionar el botón “generar estadísticas” sin escoger la variable a analizar en el caso del módulo de análisis exploratorio y al presionar el botón “ajustar semivariograma” en el caso del módulo de análisis estructural.



Figura 43. Mensaje de error generado por el sistema.

El mensaje de error mostrado en la figura 43 se generó cuando se omitió la selección de la variable y se presionaron los botones “ajustar semivariograma” o “interpolación”.

En los módulos de análisis estructural y predicción se generaron mensajes de error (figura 44), en los cuales se indicaba que los parámetros de ajuste del semivariograma no debían ser campos vacíos.



Figura 44. Mensaje de error generado por el sistema.

7.2 ANÁLISIS GEOESTADÍSTICO

A continuación se presentan los resultados obtenidos al utilizar la herramienta SIG para el procesamiento y análisis geoestadístico de las muestras de propiedades químicas de suelo en el sector río Fraile- río Amaime.

7.2.1 Análisis exploratorio de Datos

Mediante el módulo de análisis exploratorio de datos se estimaron para cada una de las variables y profundidades seleccionadas las correspondientes estadísticas descriptivas para identificar si los datos presentaban o no una distribución normal.

Las tablas 7 y 8, muestran los datos estadísticos obtenidos para cada variable según la profundidad:

Tabla 7. Estadísticas Descriptivas profundidad 0- 30 cm.

ESTADÍSTICAS DESCRIPTIVAS. PROFUNDIDAD 0 - 30 cm.

Variable	Mínimo	Máximo	Media	Mediana	Asimetría	Curtosis	D. Estandar	C.V %
<i>pH</i>	5,5	9,7	7,62	7,7	-0,26	0,57	0,6	7,87
<i>Ca</i>	6,7	53,3	18,3	18	1,27	4,28	6,31	34,48
<i>Na</i>	0,06	23,4	1,83	0,68	3,79	16,45	3,39	185,24
<i>Mg</i>	0,2	28,6	10,6	9,85	0,67	-0,009	5,65	53,3
<i>K</i>	0,04	3,7	0,61	0,5	2,44	9,38	0,49	80,32
<i>CICA</i>	2	64,3	25,3	24,7	0,68	1,54	8,75	34,58
<i>CICE</i>	9,69	68	31,31	29,66	0,81	0,65	10,49	33,5

Tabla 8. Estadísticas Descriptivas profundidad 30- 50 cm.

ESTADÍSTICAS DESCRIPTIVAS. PROFUNDIDAD 30 - 50 cm.

Variable	Mínimo	Máximo	Media	Mediana	Asimetría	Curtosis	D. Estándar	C.V %
<i>pH</i>	5,5	9,5	7,87	7,9	-0,16	0,04	0,64	8,13
<i>Ca</i>	1,7	48,4	16,57	15,6	0,8	1,39	7,31	1,39
<i>Na</i>	0,1	38,3	3,07	0,9	3,01	10,67	5,44	177,2
<i>Mg</i>	0,8	29,4	11,72	11,1	0,62	0,02	6,1	52,04
<i>K</i>	0,01	25	0,6	0,3	11,42	146,48	1,76	293,33
<i>CICA</i>	3	68,9	23,11	22,5	0,96	2,05	10,67	46,17
<i>CICE</i>	5	95,5	31,74	30,62	0,8	2,4	12,1	33,12

7.2.1.1 Análisis Gráfico

Al igual que con las estadísticas básicas, los datos fueron representados de manera gráfica utilizando histogramas, diagramas de caja y diagramas de cuantiles (Q-Q plots) generados igualmente mediante el módulo de análisis exploratorio del SIG desarrollado; dichos gráficos, en conjunto con las estadísticas descriptivas básicas mencionadas anteriormente permitieron diagnosticar si las variables presentaban una distribución normal y el tratamiento al que estas debían ser sometidas para asemejar dicha distribución.

Los resultados obtenidos mediante el análisis gráfico mostraron que con excepción del pH, todas las variables presentaban un comportamiento asimétrico, además todas las variables analizadas presentaban datos atípicos, los cuales pudieron ser los causantes de la falta de asimetría de dichas variables, estos datos se hicieron particularmente evidentes para las variables Na y K, las cuales presentaban una gran cantidad de valores alejados de la mediana de dichas variables. La normalidad de los datos fue evaluada mediante el gráfico de cuantiles obtenido para cada variable; según dicho gráfico, las variables pH, Ca, C.I.Ca presentaban una distribución muy cercana a la normal, por el contrario Na y K fueron las variables que más se alejaron de dicha distribución. En la Figura 45, se muestran los gráficos estadísticos (histograma, diagrama de cajas y Q-Q plot) generados para la primer profundidad de la variable pH, los gráficos generados para la demás variables se muestran en el **anexo D**.

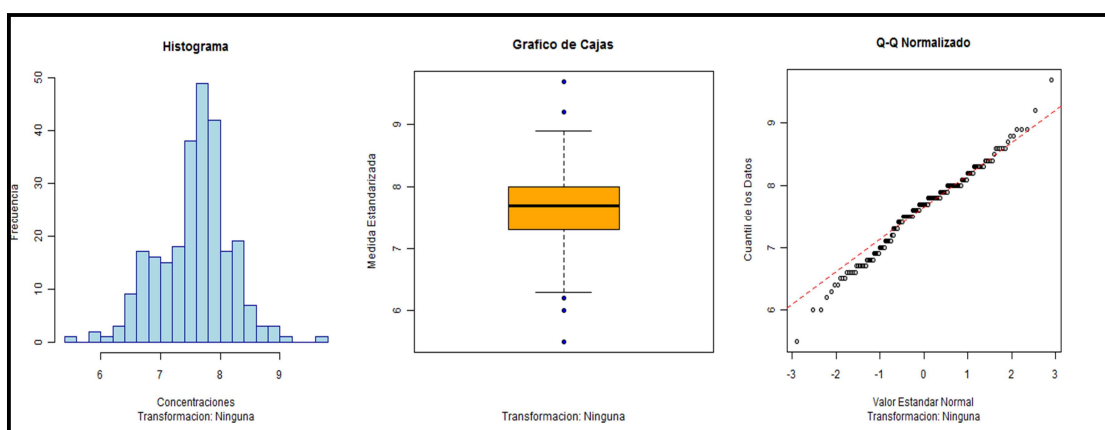


Figura 45. Gráficos estadísticos generados mediante el SIG para la variable pH. Histograma (izquierda), diagrama de cajas (centro), Q-Q plot (derecha).

Para las variables pertenecientes al primer intervalo de profundidad (Tabla 7.) fue posible observar mediante el coeficiente de variación (C.V) que el pH presentaba la menor variabilidad (C.V =7,87%), lo cual es indicador de alta homogeneidad en los datos, por su parte las variables C.I.Ca (C.V 34,58%), C.I.C.E (C.V 33,5%), Ca

(C.V 34,48%) y Mg (C.V 53,3%) presentaron una relativa heterogeneidad, mientras que el coeficiente de variación de Na (C.V 185,24%) y K (C.V 80,32%) indicaba una alta heterogeneidad.

Los datos correspondientes al segundo intervalo de profundidad (Tabla 8.) mostraron un comportamiento similar al mencionado anteriormente, con excepción del Ca cuyo coeficiente de variación en este caso C.V =1,39%, indicaba que dicha variable presenta una alta homogeneidad en esta profundidad.

En general, los resultados obtenidos se encuentran dentro de los valores esperados; esto debido a que el pH es de las propiedades químicas del suelo que menos varia, ya que es una característica intrínseca de la génesis del suelo y diversos estudios muestran que su coeficiente de variación fluctúa de 2 a 15% (Acevedo, y otros. 2008) Con respecto a elementos como Na y K presentan una alta variabilidad debido probablemente a que son altamente dependientes de las condiciones de manejo de suelo, como la incorporación de residuos orgánicos y la adición de fertilizantes (Acevedo, y otros. 2008).

De acuerdo análisis de las estadísticas descriptivas y el comportamiento gráfico se encontró que la mayoría de las variables no presentaban una distribución normal, por tal motivo se realizó una transformación a los datos, con el objetivo de que estos cumplieran dicha condición. Los criterios para realización de dicha transformación se describen a continuación:

- Para las variables cuyo coeficiente de asimetría presentaba una magnitud de entre 0 y 0.5 se asumió que presentaban una distribución normal, por lo tanto no se les realizó transformación.
- Las variables que poseían un coeficiente de asimetría con una magnitud entre 0.5 y 1 se realizó una transformación de tipo Box-Cox, el parámetro λ fue estimado para cada variable.
- Las variables que poseían un coeficiente de asimetría con una magnitud mayor a 1 se realizó una transformación de tipo logarítmica.

El tipo de transformación realizada, así como el parámetro λ (en el caso de las transformaciones Box-Cox) estimado para cada las variables transformadas se muestran a continuación:

Tabla 9. Tipo de transformación empleada

Variable	Profundidad	Transformación	λ
pH	0 -30 cm.	Ninguna	No Aplica
	30 -50 cm.	Ninguna	No Aplica
Ca	0 -30 cm.	Box-Cox	0,1
	30 -50 cm.	Box-Cox	0,5
Na	0 -30 cm.	Logarítmica	No Aplica
	30 -50 cm.	Logarítmica	No aplica
Mg	0 -30 cm.	Box-Cox	0,5
	30 -50 cm.	Box-Cox	0,5
K	0 -30 cm.	Logarítmica	No Aplica
	30 -50 cm.	Logarítmica	No aplica
C.I.Ca	0 -30 cm.	Box-Cox	0,5
	30 -50 cm.	Box-Cox	0,4
C.I.C.E	0 -30 cm.	Box-Cox	0,1
	30 -50 cm.	Box-Cox	0,6

Las estadísticas resultantes al aplicar la transformación, se presentan en las siguientes tablas:

Tabla 10. Estadísticas descriptivas con transformación. Profundidad 0 – 30 cm.

ESTADÍSTICAS DESCRIPTIVAS. PROFUNDIDAD 0 - 30 cm.

Variable	Mínimo	Máximo	Media	Mediana	Asimetría	Curtosis	D. Estándar	C.V %
Ca	2,09	4,88	3,31	3,35	-0,085	0,5	0,45	13,59
Na	-2,81	3,15	-0,22	-0,38	0,76	0,4	1,15	-522,72
Mg	-1,48	3,98	2,46	2,57	-1,26	3,32	0,82	33,33
K	-3,21	1,3	-0,76	-0,69	-0,18	-0,009	0,75	-98,68
CICA	1,46	14,6	7,36	7,49	0,13	0,31	2,21	30,02
CICE	2,54	5,24	4,04	4,04	0,0064	-0,14	0,47	11,63

Tabla 11. Estadísticas descriptivas con transformación. Profundidad 30– 50 cm.

ESTADÍSTICAS DESCRIPTIVAS. PROFUNDIDAD 30 - 50 cm.

Variable	Mínimo	Máximo	Media	Mediana	Asimetría	Curtosis	D. Estándar	C.V %
Ca	0,49	2,29	1,83	1,87	-1,6	4,47	0,24	13,11
Na	-2,3	3,64	0,13	-0,11	0,7	-0,26	1,31	1007,69
Mg	0,21	8,84	4,6	4,66	-0,05	-0,33	1,83	39,78
K	-4,6	3,21	-1,2	-1,2	0,26	1,71	1,02	-85
CICA	1,37	11,09	6,05	6,19	-0,029	0,21	1,63	26,94
CICE	2,71	24,02	11,37	11,32	0,11	0,92	3,04	26,73

Las tablas 10 y 11 sugieren que estadísticamente, las variables analizadas se asemejaron a una distribución normal. Los coeficientes de variación obtenidos para las variables K y Na fueron excluidos del análisis, esto debido en primer lugar a que los valores negativos de las medias aritméticas de tales variables impiden que se pueda hacer una interpretación de dicha medida y en segundo lugar tal coeficiente pierde significado cuando la media aritmética tiende a cero, generando coeficientes de variación altos que no necesariamente son indicadores de alta dispersión.

7.2.2 Análisis Estructural de Datos

Una vez que estadísticamente el comportamiento de los datos se asemejó a una distribución normal, se obtuvieron los semivariogramas omnidireccionales, dicho esto, no se consideró un comportamiento anisotrópico en la variable analizada ya que según este fenómeno no existe una evidencia del comportamiento de acuerdo a la dirección en que se analice la variable química del suelo. Dichos semivariogramas indicaron la presencia de estructuras de dependencia espacial, según (Giraldo Henao, 2008) esto se debe a que el valor de la semivarianza no fue constante en función de la distancia.

Los semivariogramas obtenidos fueron ajustados visualmente a los modelos teóricos esférico, exponencial, gaussiano y lineal según la forma y comportamiento de dichos semivariogramas lo permitiese. La selección del mejor modelo teórico se realizó con base en el menor error medio cuadrático (EMC) y el mejor coeficiente de determinación (r^2), este último fue el resultante del proceso de validación cruzada realizado a cada uno de los modelos ajustados. La tabla 12 muestra el modelo teórico seleccionado para el ajuste del semivariograma experimental de cada variable así como sus correspondientes parámetros.

El grado de dependencia espacial para cada una de las variables se definió según lo establecido en (Kemerer y Melchiori, 2004) quienes caracterizaron dicha dependencia según el valor de la relación pepita/meseta (C), siendo tal dependencia baja para $C \geq 0.75$, moderada para $0.25 \leq C \leq 0.75$ y alta cuando $C \leq 0.25$. Teniendo en cuenta lo anterior y al observar la tabla 12, es posible concluir que para la profundidad 1 (0-30 cm) las variables Calcio y CICA presentaron alta dependencia espacial ($C=0.18$ y $C=0$, respectivamente), las demás variables presentaron moderada dependencia espacial (los valores de C variaban desde 0.44 hasta 0.67), en cuanto a la segunda profundidad (30-50 cm) únicamente el Calcio presentó una alta dependencia espacial ($C=0.15$), con excepción de CICE y Magnesio que presentaron una baja dependencia espacial ($C=0.85$ y $C=0.78$ respectivamente) las demás variables presentaron moderada dependencia espacial (C entre 0.50 y 0.71).

Cabe destacar que en el caso de las variables que presentaron baja y moderada dependencia espacial, el valor de C nunca estuvo por debajo de 0.50, lo cual según (Giraldo Henao, 2008) podría dar lugar a predicciones imprecisas esto debido a que según dicho autor, para un modelo que describa bien la realidad el valor de C no debe ser superior a 0.50. Lo anterior lejos de interpretarse como error en los datos o durante la estimación y ajuste de los semivariogramas fue explicado según (Liu y otros, 2004) en donde se asocia la dependencia espacial de propiedades químicas con factores extrínsecos del suelo (prácticas de manejo, fertilización, etc.).

Tabla 12. Parámetros de los semivariogramas teóricos

Variable	Prof. (cm)	Modelo	Pepita	Meseta a	Rango	$C=C_0/C_0+C_1$	r^2	EMC
pH	0-30	Gaussiano	0,248	0,376	1520,927	0,66	0,31	0,572
	30-50	Exponencial	0,167	0,424	1571,945	0,39	0,25	0,547
Calcio	0-30	Exponencial	0,019	0,107	802,151	0,18	0,2	0,305
	30-50	Gaussiano	0,449	3,004	219,208	0,15	0,12	1,724
Magnesio	0-30	Esférico	2,306	3,423	6768,953	0,67	0,18	1,61
	30-50	Lineal	2,505	3,223	6210,784	0,78	0,18	1,627
Sodio	0-30	Exponencial	0,848	1,935	9152,658	0,44	0,23	1,004
	30-50	Esférico	0,77	1,088	8637,224	0,71	0,43	0,983
Potasio	0-30	Exponencial	0,386	0,59	4853,868	0,65	0,24	0,651
	30-50	Lineal	0,735	1,183	8279,107	0,62	0,28	0,862
CICA	0-30	Exponencial	0	2,956	468,932	0,00	0,2	1,646
	30-50	Exponencial	2,117	4,261	560,581	0,50	0,2	1,958
CICE	0-30	Exponencial	0,126	0,203	2244,765	0,62	0,27	0,397
	30-50	Lineal	6,871	8,097	9295,407	0,85	0,2	2,681

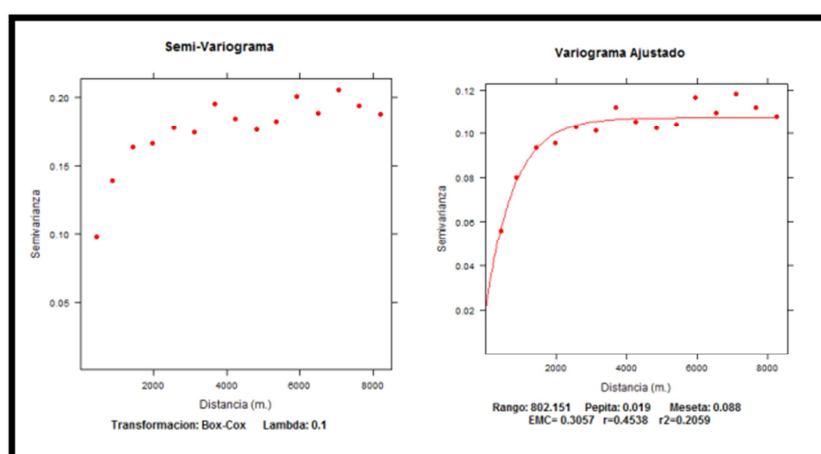


Figura 46. Semivariograma experimental (izquierda) y ajustado (derecha).

La Figura 46 muestra el semivariograma experimental y su respectivo ajuste para la variable Calcio, los semivariogramas y ajustes obtenidos para las variables restantes se muestran en el **Anexo E**.

7.2.3 Predicción Espacial

Mediante el módulo de predicción incorporado al SIG y utilizando los parámetros de los semivariogramas ajustados obtenidos anteriormente, se generaron los mapas de predicción y varianza de error para cada una de las variables analizadas en el área de estudio (**Anexo F**) utilizando Kriging Ordinario, los resultados obtenidos se muestran a continuación:

La variable pH presentó valores entre 7 y 8.4 en la primer profundidad (0-30 cm) y 6.8 a 8.8 para la segunda profundidad (30-50 cm). La superficie obtenida por medio de Kriging indicaba un comportamiento similar entre las profundidades estudiadas; los valores más altos estaban ubicados en la región central de la zona de estudio, pequeños focos con valores bajos ubicados en las regiones occidental, central y sur, se observaron también valores medios (7.6-7.8) dispersos por las áreas restantes de la zona de estudio.

La variable Calcio (Ca) presento valores que oscilaban entre los 8 a 40 y 6 a 36 para la primer y segunda profundidad respectivamente. La superficie obtenida para la primer profundidad indicaba un predominio de valores bajos (entre 8 y 20) prácticamente en toda la zona de estudio, solo se presentaron valores más altos en pequeños focos ubicados al sur y noroccidente de dicha zona. Para la segunda profundidad por su parte no fue posible establecer o determinar un comportamiento claro; en la mayor parte de la zona de estudio se presentaron valores cercanos a 16, mientras que los valores altos y bajos se ubicaban de manera dispersa y en forma de pequeñas áreas alrededor de dicha zona coincidiendo aproximadamente con las coordenadas de los puntos muestreados, lo anterior pudo deberse probablemente a la ausencia de autocorrelación espacial de la variable Calcio en dicha profundidad.

La variable sodio (Na) alcanzó valores entre 0.4 y 3.2 para la primera profundidad estudiada y de 0.8 a 8 en el caso de la segunda profundidad. Espacialmente, esta variable presento un comportamiento similar en las profundidades estudiadas, esto debido a que los valores más altos se encontraban cerca del centro de la zona de estudio y disminuían hacia las direcciones suroriental, suroccidental y nororiental hasta llegar a los límites de la zona de estudio donde predominaban los valores más bajos.

Los valores obtenidos para la variable Potasio (K) indican que estos se encontraban entre 0.24 y 1.04 para la primera profundidad y 0.2 y 2 para la segunda. Las superficies obtenidas no sugirieron la existencia de ningún tipo de

relación acerca del comportamiento de la variable en ambas profundidades. En el caso de la primer profundidad los valores más altos se encontraron en la parte superior de la región central de la zona de estudio, estos decrecían en las direcciones sureste y suroeste (esta última con un decrecimiento más notorio) hasta llegar al límite suroeste donde se encontraban los valores más bajos. En cuanto a la segunda profundidad, la zona de estudio únicamente presentaba un pequeño foco en el sector izquierdo de la región central donde se encontraron los valores más altos. La superficie obtenida no indicaba una dirección clara en cuanto al decrecimiento de los valores de dicha variable, sin embargo, fue posible observar un predominio de valores bajos en sectores de las regiones sur, nororiental de la zona de estudio.

El Magnesio (Mg) presento valores entre 5-16 y 6-16 para la primer y segunda profundidad respectivamente. El comportamiento de la variable en las profundidades estudiadas fue similar, con valores altos en la dirección noroeste-sureste y decreciendo hacia el noreste y suroeste, la única diferencia en dicho comportamiento se presentó en cercanías a la región central de la zona de estudio debido a la presencia de valores bajos en un pequeño sector de dicha región.

La variable ClCa presento valores entre 6-60 y 12-44 para la primer y segunda profundidad respectivamente. El comportamiento de dicha variable fue similar en ambas profundidades, con predominio de valores medios en gran parte de la zona de estudio, valores altos en forma de pequeños focos al noroeste y sur, valores bajos al noreste y en menor medida al sur de la zona de estudio.

La variable CICE presento valores entre 22 a 42 para la primera profundidad y 12 a 52 en el caso de la segunda profundidad. La superficie obtenida para la profundidad indicó la presencia de valores altos a lo largo de la dirección noroeste-sur, los cuales decrecían hasta llegar a las regiones noreste, sur y un pequeño sector de la región occidental con presencia de valores bajos. Para la segunda profundidad, la superficie obtenida indicaba el predominio de valores altos en toda la zona de estudio con excepción de valores bajos al noreste y suroeste, los valores más altos en este caso se encontraron al noroeste y sureste.

7.2.4 Validación de resultados

Se validaron los resultados generados por el módulo de predicción espacial utilizando la extensión Geostatistical Analyst perteneciente al *software ArcGIS*²⁹ *version 9.3* para la realizar la interpolación a las datos analizados, tal validación fue llevada a cabo teniendo en cuenta los parámetros (tipo de Kriging, modelo teórico, meseta, pepita, rango) con los que se realizaron las interpolaciones el en

²⁹ ArcGIS es el nombre de un conjunto de productos de software en el campo de los Sistemas de Información Geográfica o SIG. Producido y comercializado por ESRI. <http://www.esri.com/software/arcgis/index.html>

SIG. Lo anterior con el objetivo de comparar los resultados obtenidos en el SIG en cuanto a la veracidad de las interpolaciones y tiempos de procesamiento.

La validación realizada consistió en seleccionar de manera aleatoria dos variables, realizar la interpolación en *ArcGIS* 9.3 y comparar visualmente los resultados obtenidos

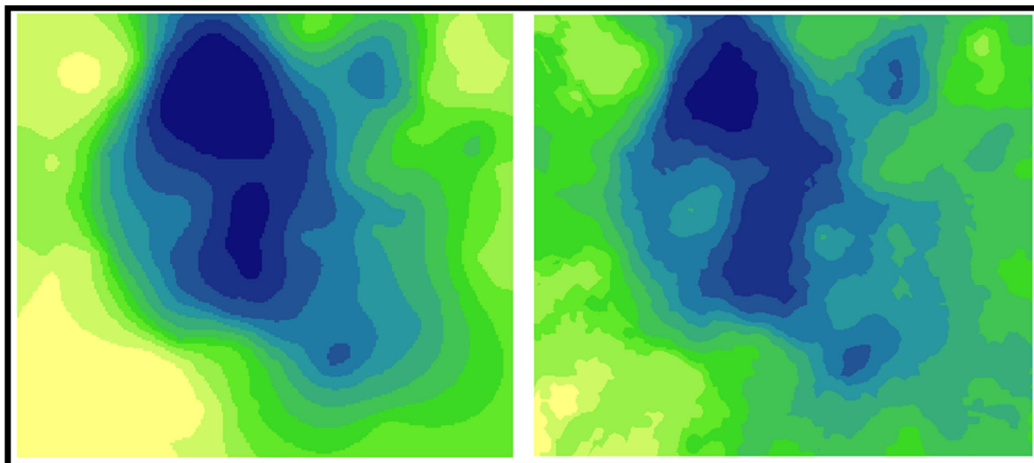


Figura 47. Comparación de superficies obtenidas para la variable Potasio. SIG-Web (izquierda), ArcGIS (derecha).

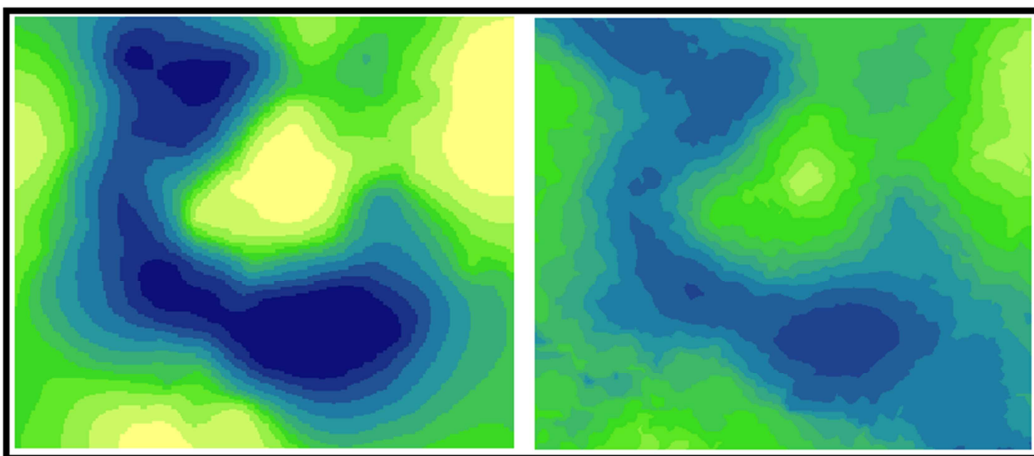


Figura 48. Comparación de superficies obtenidas para la variable Magnesio. SIG-Web (izquierda), ArcGIS (derecha).

Las figuras 47 y 48 indicaron que si bien las superficies obtenidas mediante ambas aplicaciones no fueron exactamente iguales si son muy similares, lo cual permitió establecer que las interpolaciones realizadas mediante el aplicativo son tan validas como las que se pueden obtener en cualquier *software* SIG de escritorio.

Se compararon igualmente los tiempos de procesamiento empleados por ambas herramientas para realizar las interpolaciones, se encontró que ArcGIS realizaba dichas interpolaciones de manera casi instantáneas, con tiempos inferiores a dos segundos. Los tiempos de procesamiento empleados por el aplicativo fueron probados en dos equipos de escritorio con características de hardware diferentes (tablas 13 y 14), encontrando que en el primer equipo dichos tiempos oscilaron entre 25 y 30 segundos, mientras que para el segundo los tiempos empleados variaron entre 15 y 18 segundos. Lo anterior indicó que el uso de un equipo con características de hardware más sofisticadas como un servidor reduciría notablemente estos tiempos de procesamiento.

Tabla 13. Características del equipo 1.

Equipo 1	
Sistema operativo	Windows 7
Arquitectura	32 bits
Procesador	AMD Athlon II Dual-Core 2.3 Ghz
Memoria RAM	2 GB

Tabla 14. Características del equipo 2.

Equipo 2	
Sistema operativo	Windows 7
Arquitectura	64 bits
Procesador	Intel Core I3 Ghz
Memoria RAM	6 GB

8. CONCLUSIONES

Se implementó un aplicativo web mediante la integración de herramientas SIG y estadísticas de código abierto que permite llevar a cabo un análisis geoestadístico a cualquier tipo de variable espacial con valor cuantitativo, la cual podrá ser cargada por el usuario dando la posibilidad de acceder a información georreferenciada, además de usar y reusar datos sin la necesidad de disponer de un *software especializado* instalado en un equipo y garantizando la integridad y seguridad de su información.

El aplicativo desarrollado posee una interfaz gráfica de usuario, intuitiva, que permite a los usuarios no expertos en este tipo de aplicaciones aplicar el mejor tratamiento posible a la variable analizada. Dicha interfaz permite también la clara diferenciación entre las etapas más importantes que componen un análisis geoestadístico.

La integración entre PostgreSQL y R *software* permitió el aprovechamiento de las ventajas de este último, materializándose así un servidor de almacenamiento de datos centralizado y robusto respecto a funciones de procedimiento con características complejas que permitieron proveer al aplicativo la capacidad de análisis geoestadístico.

Con excepción del pH, ninguna de las variables analizadas presentó una distribución normal al analizar los datos originales; para dichas variables fue necesario realizar transformaciones a tales datos, con el fin de que su comportamiento se asemejara al de dicha distribución.

Los semivariogramas obtenidos indicaron que las variables analizadas presentaron estructuras de dependencia espacial, ya que el valor de la meseta fue creciente en función de la distancia. Además la continuidad espacial en todas propiedades químicas reveló isotropía geoestadística en el área de estudio, este hecho implicó que solo se necesitase de un variograma para modelar el comportamiento de las variables, pero esto no quiere decir que este fenómeno siempre se vaya a presentar.

Las superficies obtenidas indican que, espacialmente, variables como pH, Na, Mg, ClCa tuvieron comportamientos similares en las profundidades analizadas, las demás presentaron comportamientos muy diferentes en dichas profundidades y en el caso de Ca en su segunda profundidad la superficie obtenida no presento un comportamiento interpretable puesto que los valores altos y bajos se situaban aproximadamente en las coordenadas de los puntos de muestreo y el resto de tal superficie parecía tener un valor constante, sugiriendo una posible falta de

autocorrelación espacial.

La validación de los resultados obtenidos señaló que las interpolaciones fueron tan veraces como las realizadas por cualquier *software* GIS de escritorio, por otra parte, en cuanto a los tiempos de procesamiento se encontró que estos fueron linealmente dependientes de las características de hardware del equipo utilizado, por lo cual era necesario disponer de equipos con características de hardware avanzadas para que los tiempos de procesamiento fueran óptimos.

9. RECOMENDACIONES

La herramienta SIG construida aunque provee un análisis espacial complejo como el geoestadístico, en un gran conjunto de campos de las ciencias aplicadas, su uso depende de unos supuestos como lo son la experiencia en el manejo de la variable que se está estudiando y un amplio conocimiento en conceptos teóricos de carácter geoestadístico. Por lo tanto para obtener buenos resultados se hace necesario conocer fundamentos sobre el sentido de la geoestadística.

Para el uso de la herramienta es conveniente tener en cuenta tanto la escala como la cantidad de puntos a procesar, ya que el análisis de áreas grandes con una red de puntos poco densa ocasiona que la predicción enmascare variaciones a pequeña escala. Debe existir un balance entre dimensiones del área de estudio y numero de datos. La selección del tamaño de dichas áreas debe ser tal que permita encontrar un semivariograma que describa un modelo de continuidad espacial que refleje los fenómenos a pequeña escala.

10. BIBLIOGRAFÍA

Acevedo, D. Cristóbal, M. E. Álvarez Sánchez, E. Hernández Acosta, R. Maldonado Torres, M. Pérez Grajales, y R. Castro Brindis. «Variabilidad Espacial de Propiedades Químicas del Suelo y su Uso en el Diseño de Experimentos.» *TERRA Latinoamericana*, Vol. 26, Núm. 4, 2008: 317-324.

Buol S.W., Hole F.D., McCracken R.J. « Génesis y Clasificación de Suelos». Ed.Trillas. México. (1983): 417.

Burrough. «Principles of Geographic Information Systems for Land Resource Assessment.», Monographs on Soil and Resources Survey No. 12. 1986, Oxford Science Publications, New York.

Costa Campos, Milton César , José Marques Júnior, Gener Tadeu Pereira, Rafael Montanari, y Diego Silva Siqueira. «Variabilidade espacial da textura de solos de diferentes materiais de origem em Pereira Barreto, SP.» *Revista Ciência Agronômica* 38, nº 2 (2007): 149-157.

Doran, J.W. and Parkin, T.B. «Defining and assessing soil quality», In J.W. Doran, D.C. Coleman, D.F. Bezdicek and B.A. Stewart (eds) Special publication 35. Madison, WI. U.S.A. (1994): 3-21.

Diaz – Frances, E «Introducción a conceptos básicos de geoestadística» Memoria seminario de estadística y medioambiente, 1993, Centro de investigación en matemáticas, Guanajuato México.

Giraldo Henao, Ramón. *Introduccion a a Geoestadistica*. Universidad Nacional,2008.

Hengl, Tomislav. *A Practical Guide to Geostatistical Mapping*.2009.

Henríquez, Carlos, Randy Killorn, Floria Bertsch, y Freddy Sancho. «La Geostatística en el estudio de la variación espacial de la fertilidad del suelo mediante el uso del interpolador Kriging.» *Agronomía Costarricense* 29, nº 2 (2005): 73-81.

Isaaks & Srivatava, (1989) «An introduction to applied geostatistics.» Oxford.

Kemerer, Alejandra C., y Ricardo J. Melchiori. «Variabilidad espacial en suelos agrícolas de la provincia de Entre Rios.». 2004.

Menezes de Souza, Zigomar, Domingos Guilherme Pellegrino Cerri, José Marcelo Colet, Henrique Luiz Antunes Rodrigues, Sérgio Paulo Graziano Magalhães, y Junqueira Rafael Araújo Mandoni. «Análise dos atributos do solo e da produtividade da cultura de cana-de-açúcar com o uso da geoestatística e árvore de decisão.» *Ciencia Rural* 40, nº 4 (2010): 840-847.

Muñoz, Juan D., Luis J. Martínez, y Ramón Giraldo. «Variabilidad espacial de propiedades edáficas y su relación con el rendimiento en un cultivo de papa (*Solanum tuberosum* L.).» *Agronomía Colombiana* 24, nº 2 (2006): 355-366.

Obando Moncayo, Franco Humberto, Andrés Mauricio Villegas Hincapié, Jorge H. Betancur P, y Leyder Echeverri Tafur. «Variabilidad Espacial de Propiedades Químicas y Físicas en un Typic Udivitrands, Arenoso de la Región Andina Central Colombiana.» *Rev.Fac.Nal.Agr* 59, nº 1 (2006): 3217-3235.

Oliver, M.A.; Webster, R. (1990): «Kriging: a method of interpolation for geographical information systems». *International Journal of Geographic Information Systems*, 4(3): 313-332.

PostgreSQL home page proyect: <http://www.postgresql.org/>

Postgis home page proyect: <http://postgis.refractory.net/>

R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>

Rossi, R. E., D. J. Mulla, A. G. Journel (1992). «Geostatistical tools for modelling and interpreting ecological spatial dependence» *Ecol. Monogr.* 62: 277–314.

Tomlin. (1990), «Geographic information systems and cartographic modelling», Prentice Hall.

Valbuena, César Augusto, Luis Joel Martinez, y Ramón Giraldo. «Variabilidad Espacial de la Fertilidad del Suelo y el Rendimiento de un Cultivo de Mango (*Mangifera Indica* L.) Localizado en La Mesa Cundinamarca.» *Revista Brasileira de Fruticultura* Vol.30, N.4, 2008: 1146-1151.

Vásquez P., José Rafael. «Caracterización de la Variabilidad Espacial de las Propiedades Físicas y Químicas en los Suelos de la Granja Experimental de la Universidad del Magdalena. » Bogotá: Universidad Nacional de Colombia- Sede Palmira, 2009.

Villatoro, Mario, Carlos Henríquez, y Freddy Sancho. «Comparación de los

Interpoladores IDW y Kriging en la Variación Espacial de pH, Ca, CICE Y P del Suelo.» *Agronomía Costarricense* 32, nº 1 (2008): 95-105.

Webster, Richard, y Margaret A. Oliver. *Geostatistics for Environmental Scientists*. John Wiley & Sons, 2007.