



FACULTAD DE INFORMÁTICA
UNIVERSIDAD DE CIENFUEGOS “CARLOS RAFAEL RODRÍGUEZ”
CARRERA DE INGENIERÍA INFORMÁTICA

TESIS DE DIPLOMA

PARA OPTAR POR EL TÍTULO DE INGENIERÍA EN INFORMÁTICA

INTERPOLACIÓN DE DATOS ESPACIALES AMBIENTALES

Autor: **Burt Collymore**

Tutores: **Dr. Roberto Henry Herrera Marrero**
MSc. Alain Muñoz Caravaca
Ing. Minerva Sánchez Llull

ENERO DE 2010

Agradecimientos

Esta es la parte más difícil de escribir en un trabajo de este tipo, al menos yo lo considero así, en primer lugar; porque se corre el riesgo de omitir a alguien que por insignificante que haya parecido, colaboró, aportó, fue fundamental, decisivo y que en un momento determinado brindó su ayuda desinteresada, en segundo lugar porque el espacio nunca alcanza para expresar como se quisiera la gratitud a todos aquellos que dieron la mano a lo largo del difícil y enrevesado camino, por eso trataré de que nadie quede olvidado, trataré de que nadie se sienta disminuido y por sobre todas las cosas que nadie deje de percibir el infinito amor que pongo en estas líneas que hoy no salen de mi pensamiento sino de mi corazón.

*Primero quiero agradecer a **Nidia Maria Lobaina** y el resto de la familia **Angel, Lioni y Eddy** por recibirme como parte, por ayudarme y aguantar todas mis locuras en el largo camino de ser ingeniero.*

*Agradezco a mis tutores **Dr. Roberto Henry Herrera Marrero, MSc. Alain Muñoz Caravaca** y **Ing. Minerva Sánchez Llull** por su orientación profesional, por la oportunidad de introducirme en este tema de investigación y por ser buenas personas.*

*Al **MSc. Miguel Santana** por ser un buen amigo y por su dedicación en ayudarme a terminar la fase final de la tesis:- gracias por introducirme a programar en **ETEX**.*

*A la **Ing. Lizette Montero** por asegurarse que el currículo ajustado terminara con éxito y por su presencia constante cuando necesitaba ayuda y consejo.*

*Al **Ing. Juan Felipe Medina** y **MSc. Daimarelys Acevedo Cardoso** por aclarar mis dudas.*

*Al **MSc. Oscar José Alejo Machado** por sus consejos.*

A las personas del departamento de Informática por compartirme sus conocimientos.

A las personas del Centro de Estudios Ambientales de Cienfuegos (CEAC) donde realicé el trabajo, y a mis compañeros del grupo de Informática por compartir conmigo la experiencia de ser universitario.

BURT COLLYMORE.
Cienfuegos, Enero de 2010.

RESUMEN

En el presente trabajo se realizó una investigación sobre la representación espacial de los parámetros de control ambiental medidos en el Centro de Estudio Ambientales de Cienfuegos, Cuba (CEAC), para el monitoreo y caracterización de la Bahía de Cienfuegos, con el propósito de identificar cuáles proporcionan una mejor descripción de los procesos en estudio.

Los programas profesionales empleados en el CEAC para la interpolación de datos espaciales proponen algoritmos que en muchos casos no satisfacen las necesidades de los investigadores y presentan además la dificultad de no poder operar sobre los algoritmos implementados en ellos. Estas limitaciones influyen negativamente en el proceso de selección de las técnicas adecuadas para cada caso particular. Se identificó también la necesidad de una descripción adecuada de los algoritmos así como de las técnicas de validación matemática que aporten una métrica de calidad del resultado final en correspondencia con los valores reales medidos en campañas de investigación. En esta tesis se concluye que las técnicas de interpolación basadas en Kriging y Splines conducen a las mejores estimaciones validadas por las técnicas de “Cross-Validation” y “Jackknife”.

La investigación concluye con el desarrollo de un software, sobre la plataforma MATLAB, que permite el preprocesamiento de los datos, las aplicaciones de técnicas de interpolación y la visualización de los resultados en un ambiente gráfico, con todas las potencialidades de los Software profesionales y cuenta con la correspondiente ayuda que ejemplifica varias situaciones reales, descriptivas de los problemas más comunes. El programa permite, además, la asimilación de tablas de datos de Excel que es la forma habitual donde los investigadores almacenan sus mediciones de campo. Con esta solución se logra un aporte metodológico y práctico en esta rama del conocimiento.

ÍNDICE

	Pág.
RESUMEN	I
ÍNDICE	II
INTRODUCCIÓN	1
1. Estado actual de la interpolación de datos espaciales	7
1.1. La importancia del tratamiento de datos ambientales	7
1.2. ¿Por qué interpolar?	8
1.3. ¿Cómo se generan los datos?	10
1.3.1. Datos primarios y secundarios	11
1.3.2. Simulación de datos espaciales	11
1.3.3. Generación de datos espaciales a través de Muestreos (Sampling)	12
1.4. Métodos de Interpolación	14
1.4.1. Nearest Neighbor (NN)	15
1.4.2. Natural Neighbor Interpolation (NNI)	16
1.4.3. Inverse distance weighted method (IDW)	18
1.4.4. Interpolación basada en Splines	19
1.4.5. Interpolación basada en Kriging	19
1.5. Métodos de Validación	22
1.5.1. Comparación de los métodos de interpolación	24
1.6. Conclusiones parciales	25

2. Materiales y Métodos	26
2.1. Interpolación basada en Splines	26
2.1.1. Teoría	26
2.1.2. Formulación Matemática: Splines Cúbicos	27
2.1.3. El cálculo de un Spline Cúbico Natural	29
2.2. Interpolación basada en Kriging	32
2.2.1. Teoría	32
2.2.2. Formulación Matemática	36
2.3. Conclusiones parciales	40
3. Métodos de diseño y modelación del Software	42
3.1. Introducción	42
3.2. Herramientas y Principios	42
3.2.1. La tecnología y lenguaje de programación en MATLAB	42
3.2.2. Conceptos de programación orientada a objetos (POO)	43
3.2.3. La programación en COM y controles de Active X	43
3.2.4. Ejecutables de MATLAB (MEX-Executables) creado con Fortran, C/ C++	44
3.2.5. Modelación en UML (Unified Modeling Language)	44
3.2.6. La metodología RUP (Rational Unified Process)	45
3.2.7. Rational Rose 2003	46
3.2.8. Adobe Photoshop CS3	46
3.2.9. Microsoft Excel	46
3.2.10. Borland C++ Builder 2007	46
3.3. Modelado de diferentes vistas de una Arquitectura	47
3.4. Modelo de Dominio	49
3.4.1. Representación del Modelo de Dominio	49
3.4.2. Diagrama de Clases de Dominio	50
3.4.3. Descripción del Modelo de Dominio	51
3.4.4. Reglas del dominio a considerar	51
3.5. Modelo de Sistema	52
3.5.1. Representación del Modelo de Sistema	52
3.5.2. Modelo de casos de uso del sistema	56

3.6. Construcción del sistema	67
3.6.1. Modelo de Diseño	67
3.6.2. Diagrama de implementación	68
3.7. Conclusiones parciales	69
4. Análisis y Discusión de los Resultados	70
4.1. Caracterización de la muestra	70
4.2. Caracterización del software	72
4.2.1. ArcGIS9	72
4.2.2. SURFER8	73
4.2.3. IDEA	74
4.3. Comparación de los métodos de interpolación en ArcGIS9, SURFER8 e IDEA . .	76
4.3.1. Interpolación con ArcGIS9	76
4.3.2. Interpolación con SURFER8	77
4.3.3. Interpolación con IDEA	78
4.4. Validación de los métodos de interpolación	79
4.4.1. Estadística de la validación en ArcGIS9	80
4.4.2. Estadística de la validación en SURFER8	80
4.4.3. Estadística de la validación en IDEA	83
4.5. Comparación del error de medición	84
4.6. Costo computacional basado en la arquitectura	85
4.7. Conclusiones parciales	87
CONCLUSIONES	89
RECOMENDACIONES	90
REFERENCIAS BIBLIOGRÁFICAS	91
Glosario	94
A. Prototipos de los Casos de Uso del sistema	96
A.1. Caso de Uso: Entrar datos espaciales	96
A.2. Caso de Uso: Representar gráficamente los datos crudos	98

A.3. Caso de Uso: Representar gráficamente las etiquetas	99
A.4. Caso de Uso: Interpolar datos	100
A.5. Caso de Uso: Representar gráficamente los datos interpolados	101
A.6. Caso de Uso: Representar etiquetas de datos interpolados	102
A.7. Caso de Uso: Validar los datos interpolados	103
A.8. Caso de Uso: Generar reporte estadístico general	104
A.9. Caso de Uso: Generar reporte estadístico comparativo	105
A.10.Caso de Uso: Consultar ayuda	106
A.11.Caso de Uso: Comparar métodos de interpolación	107
A.12.Caso de Uso: Representar gráficamente la comparación de interpoladores	108
A.13.Caso de Uso: Crear modelo de variograma	109
A.14.Caso de Uso: Representar gráficamente el variograma	110
A.15.Caso de Uso: Crear nubes de variograma	111
A.16.Caso de Uso: Realiza proceso de clúster	112
A.17.Caso de Uso: Crear máscara	114
A.18.Caso de Uso: Exportar datos interpolados	115
B.	116
B.1. Mex Executables (ficheros ejecutables de MATLAB)	116

LISTA DE TABLAS

4.1. Estadísticos de las mediciones de salinidad	71
4.2. El resultado de validar los métodos de interpolación en ArcGIS9	80
4.3. El resultado de validar los métodos de interpolación en SURFER8	82
4.4. El resultado de validar los métodos de interpolación en IDEA	84
4.5. Comparación del costo computacional en SURFER8 e IDEA	86

LISTA DE FIGURAS

1.1. Tipos de muestreo espacial.	14
1.2. Diagrama de Voronoi, tomada de MATLAB R2007b.	16
1.3. Polígonos de Thiessen.	17
1.4. Las coordenadas de los vecinos naturales de x	18
1.5. El modelo semi-variograma	20
2.1. Splines Cúbicos.	27
2.2. El muestreo hipotético y situación de estimación.	34
2.3. Vector de separación h entre dos puntos.	34
2.4. Una estimación más realista - el valor del bloque es requerido.	38
3.1. Vista general de los elementos de UML.	45
3.2. Modelado de la arquitectura de un sistema.	48
3.3. Conjunto de clases relacionadas en el Modelo de Dominio.	50
3.4. Actores del Modelo de sistema.	56
3.5. Diagrama de casos de usos divididos por Paquetes.	58
3.6. Diagrama de los casos de usos de Interpolación.	58
3.7. Diagrama de los casos de usos de Visualización.	58
3.8. Diagrama de los casos de usos de Validación.	59
3.9. Diagrama de los casos de usos generales.	59
3.10. Representación del diagrama de clases del Diseño.	68
3.11. Representación del diagrama de componente de implementación.	69
4.1. Distribución de las estaciones en la Bahía de Cienfuegos	71
4.2. Kriging en ArcGIS9	72
4.3. Selección del radio de búsqueda en ArcGIS9	73
4.4. Métodos de interpolación en SURFER8	74
4.5. Métodos de interpolación en IDEA	75

4.6. Mapas de interpolación de ArcGIS9	77
4.7. Mapas de interpolación de SURFER8	78
4.8. Mapas de interpolación de IDEA	79
4.9. Comparación de la interpolación en SURFER8	81
4.10. Comparación de los modelos predictivos de Crossvalidation	85
B.1. El ciclo del proceso MEX en C/C++	116

*A mi mamá Cecilia,
mi familia,
por todo su sacrificio y su amor.*

*A Dunia,
mi amor,
la única persona que podía entender mi egoísmo por mis estudios.*

*A mi hijo WilloW,
un día espero ser el padre dedicado y no ocupado.
Y a mi niña por nacer.*

INTRODUCCIÓN

Importancia del tema

Muchas disciplinas científicas usan los modelos matemáticos para describir los complejos procesos ambientales. Técnicas como la interpolación ayudan a los científicos a descubrir o a predecir las tendencias para una variable específica de interés, como los niveles de contaminación en la tierra, la polución en el aire, contenido de arsénico en el agua, patrones de distribución de lluvia, presión barométrica, la demografía, el crimen y muchas más (Webster, 2007).

En los estudios medioambientales, disciplinas como la geoestadística usan la interpolación extensivamente. Cuando los cambios en el ambiente son más evidentes, sobre todo vistos desde imágenes satelitales y fotografías aéreas, otros permanecen más sutiles, y propiedades como la temperatura o la composición química raramente pueden visualizarse, entonces, es vital confiar en las mediciones y los análisis de muestras. Los problemas involucrados en la recolección de las muestras están basados en el hecho de que las mediciones se realizan en volúmenes pequeños de material o áreas de unos centímetros hasta metros, según la escala, que pueden ser considerados como muestras puntuales (*“point samples”*) conocido técnicamente como *“supports”* (Webster, 2007). En otros casos podrían hacerse varias mediciones en áreas más extensas donde el esfuerzo involucrado y el costo asociado al proceso que harían no factible la tarea. La alternativa sería medir tales propiedades en sólo un número finito o limitado de lugares. Para superar estas dificultades de muestreo en áreas extensas, los científicos usan datos espaciales recolectados en un conjunto de puntos esparcidos y estiman o predicen la conducta de estas variables en puntos donde no se han realizado mediciones. El método aplicado garantiza una estimación que representa la realidad observada con el mínimo error.

Los métodos de interpolación usados dependen principalmente de las variables estudiadas, pero según Lloyd (2007) pueden ser divididos en dos grandes grupos: los métodos globales y locales. Estos métodos pueden distinguirse por el hecho de que los métodos globales usan todos los datos para la predicción, mientras los métodos locales normalmente usan sólo algún subconjunto de los datos para hacer cada predicción. En este trabajo se describe cada caso y las posibles aplicaciones según los requerimientos de la investigación y las variables a interpolar.

Con tantos métodos de interpolación disponibles, es crítico evaluar cuan bien ellos representan el proceso estudiado, antes de tomar una decisión acerca de cuál método es el más apropiado para el análisis de datos medioambientales espaciales. Es importante notar que las diferentes aplicaciones en las ciencias de la tierra requieren métodos distintos para la interpolación y que estos métodos actúan diferentemente en cada situación, es decir, no hay ningún método clasificado por defecto para determinados datos y en las condiciones dadas, cualquier método puede superar a otro. Por lo que se deben proveer las herramientas y dejar a criterio del investigador y los parámetros de error, encontrar cuál es el que mejor describe el proceso que se estudia. Las dificultades en estos métodos están en su demanda computacional, la complejidad o simplicidad de su naturaleza. Es por estas razones que se presenta una comparación de estos métodos de interpolación a través del uso de técnicas de validación de la calidad como “Cross-validation”, “Bootstrap” y “Jackknife”. Estas técnicas determinan hasta qué punto se minimizan los errores de interpolación, garantizando la exactitud de los datos ambientales representados.

Además de brindar detalles de la descripción teórica de los métodos de interpolación se validan los resultados con uso de datos medioambientales reales obtenidos del Centro de Estudios Ambientales de Cienfuegos (CEAC), perteneciente al Ministerio de Ciencia, Tecnología y Medio Ambiente. Se desarrolló un nuevo toolbox de MATLAB (IDEA, Interpolación de Datos Espaciales Ambientales), que constituye una herramienta de soporte en la investigación científica en este campo. Los resultados muestran que las interpolaciones basadas en Splines y Kriging proporcionan la mejor aproximación a los datos espaciales ambientales reales.

Diseño Teórico

Situación Problémica:

La Bahía de Cienfuegos es un ecosistema costero de gran atractivo y productividad biológica, con grandes recursos naturales y múltiples usos para fines económicos, comerciales, residenciales y turísticos, por lo que se presenta como una de las más importantes bahías del país. Los estudios ambientales realizados muestran una tendencia a perder la capacidad natural para su auto-depuración debido a la contaminación de los espacios terrestres y marinos producidos por las constantes emisiones atmosféricas, derrames de petróleo y contaminantes líquidos de naturaleza orgánica e inorgánica (Tolosa, Mesa-Albernas, & Alonso-Hernandez, 2009). Las fuentes que más contribuyen a la contaminación son las de tipo industrial, aunque también se considera el aporte de unidades de servicios, transporte marítimo y las producidas por los asentamientos humanos, en este caso se considera el aporte antropogénico de una ciudad de 150 000 habitantes.

La importancia de impulsar el manejo integrado en zonas costeras para la conservación de sus recursos naturales fue descrita explícitamente en la reunión celebrada en Río de Janeiro, en 1992 denominada “Cumbre de la Tierra” que en el capítulo 17 de la Agenda 21, refleja la protección y utilización racional de los recursos de los océanos, los mares y las costas (Castro-Ruz, 2007). Con ello quedó establecido el desarrollo de las políticas y los programas que debían acoger las naciones para alcanzar modelos de desarrollo socio-económicos sostenibles.

En la actualidad la dimensión ambiental se concreta en la política y los planes de gestión que desarrollan el Ministerio de Ciencia, Tecnología y Medio Ambiente, los Organismos de la Administración Central del Estado y las empresas consultoras para la protección del medio ambiente y la actividad científica como el Centro de Estudios Ambientales de Cienfuegos (CEAC), encargado del control ambiental de la Bahía de Cienfuegos. Sin embargo, la atención a los problemas socio-ambientales que provoca la contaminación ambiental, todavía no presenta resultados favorables ya que los estudios realizados para evaluar la concentración de arsénico en macroalgas de la Bahía de Cienfuegos (Castellanos *et al.*, 2005), revelan la presencia de contaminantes en concentraciones por encima de las aceptadas en la norma cubana para zonas

habitables (NC, 1987).

El CEAC tiene amplia experiencia en la gestión ambiental y en investigaciones, con una producción científica relevante a nivel internacional. Sin embargo, el tratamiento de los datos generados en las campañas de investigación sigue siendo un punto débil que pudiera ser resuelto con las herramientas matemáticas e informáticas, que permitan la extracción de información útil de las bases de datos coleccionadas por años, en este centro de investigación.

Problema Científico:

¿Cómo representar los datos espaciales a partir de mediciones reducidas en los monitoreos de variables ambientales?

Objetivo General:

Desarrollar un software para la interpolación de variables ambientales.

Objetivos Específicos:

- ▷ Diseñar un software para el tratamiento y visualización de los datos ambientales.
- ▷ Desarrollar el software Diseñado.
- ▷ Analizar los resultados del trabajo.
- ▷ Validar los resultados del trabajo.

Las **tareas** a realizar para cumplir con los objetivos propuestos son:

- ▷ Hacer un estudio sobre el tratamiento de datos ambientales espaciales.
- ▷ Implementar algoritmos para la validación de los métodos de interpolación.
- ▷ Diseñar e implementar la tecnología que permita la comparación de los métodos de interpolación.
- ▷ Diseñar e implementar la tecnología que permita la simulación de datos reales.

- ▷ Diseñar e implementar la tecnología que permita el análisis multivariado de los datos.
- ▷ Brindar la información necesaria para la posible utilización de tales tecnologías.

Justificación del estudio:

Necesidad social: El gran desarrollo social e industrial que se produce en las zonas costeras, ha provocado un incremento significativo de la contaminación ambiental a escala local. La pérdida de la calidad del ambiente en la Bahía de Cienfuegos repercute en el medio marino y terrestre ya que representa un riesgo para la comunidad biológica y la salud del hombre, además de contribuir a que ocurran cambios globales asociados a este tipo de contaminación. Por tanto el uso de este software contribuye a una mejor interpretación de los procesos ambientales, así como a una toma de decisiones científicamente fundamentada, en relación con el uso de la Bahía.

Conveniencia: En la actualidad el paquete de software MATLAB usado en el CEAC ha sido un desafío en la construcción y validación de un modelo basado en el Kriging. Esta investigación es de gran utilidad en el sentido de mejorar el desempeño tanto del uso del Kriging en el paquete de software como de la calidad de la información recibida por los científicos que lo emplean para el tratamiento de datos ambientales. También, ofrece una nueva vía de validar y comparar los métodos existentes.

Objeto de la Investigación:

Análisis exploratorio de datos.

Campo de acción:

Interpolación de datos ambientales.

Idea a Defender:

Con el desarrollo de un software para la interpolación se logra una adecuada representación espacial de los datos ambientales.

Para alcanzar esta meta se organizó el informe de la siguiente manera:

- ▷ Capítulo I: Se describe el estado actual de la interpolación de datos espaciales. Se parte de la importancia del tratamiento de datos medioambientales, las razones generales para la interpolación de datos espaciales y la forma en que se colectan los datos. En la Sección 1.3 se da una apreciación global de las distintas aproximaciones y los factores que llevan a una estrategia exitosa en el muestreo. En la sección 1.4 se discuten algunos de los métodos ampliamente usados en la interpolación; se exponen, en una breve introducción, las categorías a las cuáles pertenecen los métodos y las distinciones básicas entre ellos. Se explica el diagrama del Voronoi y los métodos de Triangulación de Delaunay, ampliamente usados en los métodos de interpolación. También se enuncian algunas de las ventajas y desventajas de estos métodos. El capítulo concluye con las métricas de validación de los métodos de interpolación propuestos.
- ▷ Capítulo II: Materiales y Métodos. Se presenta la teoría y la formulación matemática de los métodos de interpolación Splines Cúbicos y Kriging.
- ▷ Capítulo III: Ingeniería del Software. Programación orientada a objetos en MATLAB. Por primera vez en la Tesis de Ingeniería Informática de la UCF, se realiza un estudio de Ingeniería del Software basado en la plataforma MATLAB. Se hace un análisis detallado de las estrategias de la Programación Orientada a Objetos (POO) para la solución de casos reales bajo exigencias de soporte de investigación.
- ▷ Capítulo IV: Análisis y discusión de los resultados. Se presenta a través de la cross-validación un análisis del error de predicción de los métodos de interpolación de ArcGIS9, SURFER8 y IDEA (el software desarrollado para esta investigación). Se hace una caracterización de una muestra de datos de salinidad de la Bahía de Cienfuegos utilizados en el análisis. Se hace una caracterización de los programas explorando algunas de las ventajas y desventajas, y las distinciones básicas entre ellos. Se presenta la estadística de validación de la interpolación realizada en estos paquetes de software en forma comparativa. También se hizo un análisis del costo computacional basado en la arquitectura del software IDEA.

Capítulo 1

Estado actual de la interpolación de datos espaciales

La estimación de datos espaciales es vital en la ciencia medioambiental debido a las dificultades de costo y esfuerzo asociadas con el procesamiento de tales datos. Este capítulo está orientado hacia los aspectos teóricos sobre el análisis de los datos medioambientales con el objetivo de facilitar la presentación de los métodos interpoladores utilizados para su procesamiento. En este capítulo se proporcionan los resultados de reconocidas comparaciones de efectividad de cada método (Cross-validation (“leave-one-out”) y Jackknife) y se indican los métodos más apropiados.

1.1. La importancia del tratamiento de datos ambientales

La necesidad del análisis de datos espaciales medioambientales puede estar reflejada en el número de situaciones que se presentan al tratar con eventos naturales que ocurren en el espacio y el tiempo. Todos los eventos tienen coordenadas espacio-temporales vinculadas a ellos, es decir, que todos ocurren en alguna parte y en algún momento, esto no significa que en áreas como la ciencia experimental esta información sea importante porque el resultado del experimento sólo depende de la variable de interés y no de su ubicación. En contraste de la ciencia experimental, las ciencias observacionales como las ciencias sociales y medioambientales confían en el uso de datos espaciales extensivamente.

En las ciencias sociales, los eventos se estudian en distintos lugares, y un lugar particular puede tener gran influencia en el resultado de tal proceso. Anotando donde estos eventos tienen lugar, puede dar grandes beneficios debido al hecho de que ellos pueden ser referenciados en forma

cruzada o pueden unirse a otras bases de datos como el censo y la salud para proporcionar una vista más comprensiva del proceso social.

La rama de las ciencias medioambientales tiene muchas disciplinas que usan los datos espaciales. El reconocimiento de la importancia de espacio o locación en estas disciplinas se evidencia en el uso de prefijos como geográfico, espacial, medioambiental o regional: la epidemiología geográfica y medioambiental, la arqueología espacial, la criminología medioambiental, y la economía regional (Haining, 2004).

El estudio de aumentos de enfermedades infecciosas genera preguntas sobre los orígenes del brote, cómo se desarrolla a través de tiempo, la forma geográfica y magnitud de su propagación, y las condiciones bajo las cuales un brote pequeño puede convertirse en una epidemia mayor, que afecte una proporción grande de la población. Predecir el curso y la propagación geográfica de una enfermedad infecciosa, es imperante para intentar controlarla, pero cada una de las preguntas individuales genera preguntas más amplias sobre el papel del lugar y el espacio, y estos han influido en modelos matemáticos e investigaciones empíricas de enfermedades infecciosas. Por ejemplo, ciertas características de los lugares de aparición han ayudado a comprender los orígenes de una epidemia.

Estas necesidades de procesamiento de datos en las ciencias de la salud, sociales y ambientales hacen más fuerte el nexo con las ciencias de la computación. El procesamiento de estos volúmenes de datos para generar información útil a los investigadores, se deriva en conocimientos que influyen en la toma de decisiones. La aplicación de conceptos informáticos debe conducir a una exploración de los datos para brindar información entendible, precisa, confiable y en tiempo.

1.2. ¿Por qué interpolar?

El uso de la interpolación es inmenso y la opción del método depende del problema dado. Richard Webster en la página 1 de su libro (Webster, 2007) ha respondido la pregunta de la necesidad por la interpolación cuando él dio una introducción titulada ¿why Geostatistics? Él declara explícitamente: “Imagine la situación: un granjero le ha pedido que inspeccione la tierra

de su granja. En particular, él quiere que usted determine el volumen del fósforo; pero él no se satisfará con el valor del promedio por cada campo. Él quiere más detalle ahora para que él pueda agregar el fertilizante sólo donde la tierra es deficiente, no por todas partes. El estudio involucra la toma de numerosas muestras de tierra que usted debe transportar al laboratorio para el análisis. Usted seca las muestras, las procesa, extrae el fósforo con algún reactivo y finalmente lo mide en los extractos. El proceso entero consume tiempo y es costoso. ¡No obstante, al final usted tiene los datos de todos los puntos que usted tomó- justo lo que el granjero quiere, usted podría pensar!

La desilusión del granjero es evidente, sin embargo. ‘Oh’, él dice, ‘esta información es para un juego de puntos, pero yo tengo que cultivar surcos continuos de tierra. Yo realmente quiero saber cuánto fósforo la tierra contiene por todas partes. Yo comprendo que es imposible; no obstante, debe brindarme realmente un poco de información de los lugares entre sus puntos de prueba. ¿Qué usted puede decirme sobre aquéllos?’

Esto levanta problemas extensos sobre lo que usted debe pensar ahora. ¿Usted puede decir los valores esperados en los lugares intermedios entre los puntos de la muestra y encima de los bloques, o de la anchura del propagador de fertilizante del granjero. Y qué densamente usted debe muestrear para que tal información sea fiable? En todo momento usted debe considerar el equilibrio entre el costo de proporcionar la información y las ganancias financieras del granjero. En el contexto más amplio puede haber una ganancia adicional si usted puede ayudar a evitar que se sobre fertilice y con ello proteger el ambiente de la contaminación por el exceso de fósforo. Su tarea, como agrimensor es ser capaz de utilizar escasos datos disponibles para estimar o predecir, los valores medios de fósforo en la tierra”.

En los capítulos posteriores continúa explicando, “las mediciones de las componentes ambientales, la tierra, el clima y el agua, son hechas en pequeñas partes de estos materiales separados entre sí por distancias relativamente grandes. Ellos constituyen una muestra de un continuo que no puede medirse en todas partes. A las personas que hacen las mediciones o a sus clientes les gustaría casi siempre saber los valores que están en el espacio intermedio.

La interpolación también se usa en ingeniería y otras ciencias, para ello es frecuente disponer de un cierto número de puntos obtenidos por muestreo o a partir de una medición o experimento y

pretender construir una función que los ajuste. En el sub-campo matemático del análisis numérico, se denomina interpolación a la construcción de nuevos puntos partiendo del conocimiento de un conjunto discreto de puntos y un problema estrechamente ligado con el de la interpolación es la aproximación de una función complicada por una más simple. Si se tiene una función cuyo cálculo resulta costoso, se puede partir de un cierto número de sus valores e interpolar dichos datos construyendo una función más simple. En general, por supuesto, no se obtendrán los mismos valores evaluando la función obtenida que si se evaluara la función original, si bien dependiendo de las características del problema y del método de interpolación usado la ganancia en eficiencia puede compensar el error cometido.

En todo caso, se trata, a partir de n puntos distintos x_k llamados nodos de obtener una función f que verifique:

$$f(x_k) = y_k : k = \overline{1, n}$$

A $f(x)$ se denomina la función interpoladora de dichos puntos, siendo el valor del k -ésimo nodo en la función original. Algunas formas de interpolación que se utilizan con frecuencia son la interpolación lineal, la interpolación polinómica, de la cual la anterior es un caso particular, o la interpolación por medio de splines.

1.3. ¿Cómo se generan los datos?

Los datos espaciales son adquiridos de muchas maneras diferentes. Esta sección del capítulo muestra que se pueden obtener los datos espaciales desde fuentes de datos primarios, secundarios y de simulación. La mayoría de esta sección trata sobre el “Muestreo” como el medio principal de generar los datos espaciales. En su introducción sobre las formas de recoger datos Trauth (2006), describe un juego de datos como tener un tamaño de muestra muy limitado y luego como una estrategia apropiada de “Muestreo” es el primer paso para obtener un buen juego de datos.

1.3.1. Datos primarios y secundarios

Los datos primarios son recogidos por un investigador para cumplir los objetivos específicos de un proyecto. En la ciencia observacional los datos primarios se originan del trabajo del campo de estudio y de la muestra de estudio (Webster, 2007).

Los datos primarios tienen la ventaja en el sentido que sus exactitudes y sus confiabilidades dependen de los métodos empleados por el investigador en el proceso de adquirir tales datos y no por las fuentes externas como en el caso de datos secundarios.

Las fuentes de los datos espaciales secundarios incluyen los mapas, datos de censos sociales, económicos y demográficos nacionales y regionales, datos generados por los cuerpos públicos como la salud, las autoridades locales así como los juegos de datos comerciales generados por las instituciones privadas, por ejemplo, el menudeo y los sectores financieros (Haining, 2004).

Los satélites también son una fuente importante de datos medioambientales, por ejemplo, el NASA/GSFC/METI/ERSDAC/JAROS y U.S./Japan ASTER Science Team proporcionan los datos del satélite para el servicio del Sistema de Información Geográfico (GIS) (NASA, 2009). Una fuente alternativa de valiosos datos viene de Shuttle Radar Topography Mission (SRTM). Este incorpora un sistema de radar que voló junto al Transbordador Espacial “Endeavour” durante una misión de 11 días en el 2000. SRTM es un proyecto internacional manejado por la National Geospatial-Intelligence Agency (NGA) y la Administración de Aeronáutica Nacionales y del Espacio (SRTM, 2009).

1.3.2. Simulación de datos espaciales

Algunos datos se generan entrando los datos de la muestra en un modelo para generar una superficie espacial de valores de los datos. Esto puede hacerse si la variable de interés es difícil o cara de conseguir, por ejemplo, el monitoreo de contaminación del aire. Usando los métodos de la simulación, los mapas de contaminación del aire para una área específica pueden ser contruidos combinando los datos en los puntos conocidos con los datos climatológicos y asunciones sobre cómo los contaminantes se dispersan. El rendimiento generado del modelo de simulación puede

calibrarse y puede verificarse contra los datos conocidos de la muestra para generar los mapas de contaminación del aire (Haining, 2004).

1.3.3. Generación de datos espaciales a través de Muestreos (Sampling)

El propósito detrás del muestreo espacial es hacer las inferencias sobre una población dónde cada miembro tiene una referencia geográfica, en base a un subconjunto de individuos de esa población. El muestreo se usa en lugar de hacer un estudio completo por varias razones. La población puede ser tan grande que un estudio completo sería físicamente imposible o no factible (por ejemplo: estimando la longitud media de guijarros en una playa) puede haber una infinidad de locaciones dónde las medidas podrían tomarse como en el caso de calidad del aire, o profundidad de la tierra. El costo de adquirir la información sobre cada miembro en la población puede negar la posibilidad de hacer el estudio completo (Haining, 2004).

El desarrollo de una estrategia exitosa de muestreo depende de varios factores:

- ▷ Lo que será estimado: ¿Si es una propiedad espacial o no espacial de la población que es de interés primario? Un ejemplo de una propiedad no espacial de una población es el nivel promedio de un atributo en un área o la proporción de la población que excede un cierto valor del umbral. Ellos son llamados propiedades no espaciales porque el analista sólo está interesado en ‘cuánto’ no ‘dónde.’
- ▷ El tamaño de la muestra: Este parámetro incluye el volumen de la muestra o su peso, así como, el número de muestras coleccionado en el campo (Trauth, 2006). Si los datos de la población varían de una manera excesiva, entonces el tamaño de la muestra debe ser bastante grande para reflejar esa variabilidad y aumentar la exactitud; por otra parte es una regla general para dejar el tamaño de la muestra tan pequeño como sea posible para reducir tiempo y esfuerzo para analizarlo.
- ▷ El tipo de muestreo: La estrategia apropiada depende del tipo de objeto a ser analizado, el propósito de la investigación y el nivel requerido de confianza del resultado final (Trauth,

2006). Las principales estrategias de muestreo son: aleatoria, regular, aleatoria estratificado y sistemático.

- La muestra aleatoria: Un número (n) de puntos se selecciona para que cada miembro de la población tenga una oportunidad igual e independiente de ser seleccionado.
- El muestreo regular: Las muestras son basadas en una reja regular uniformemente espaciada.
- La muestra aleatoria estratificada: La población a ser probada se divide en los estratos y, dentro de cada uno de éstos; se seleccionan los puntos según el método de muestra aleatoria. Figura 1.1(a) muestra una estrategia aleatoria estratificada basada en nueve estratos con uno prueba tomado de cada estrato al azar.
- El muestreo sistemático: Figura 1.1(b) muestra un tipo de muestreo sistemático, ‘Centric Systematic Sampling’, con los estratos cuadrados y un punto tomado del centro de cada estrato. Las propiedades aleatorias (aleatorización) puede introducirse seleccionando la posición del punto al azar en el primer estrato. El resto de los ($n-1$) puntos se posicionan relativamente a la posición seleccionada para el primer punto. O sea, si el primer punto se selecciona en el centro, todos los puntos restantes se seleccionarán en el centro; si se selecciona en la esquina derecha superior, los puntos restantes se seleccionarán en la esquina derecha superior y así en lo adelante.
- Otros: Otros tipos de muestreo pueden ser apropiados en circunstancias particulares, los esquemas de muestreo como el muestreo uniforme ‘uniform sampling’, muestreo por conglomerado ‘cluster sampling’, muestreo transversal ‘traverse sampling’, muestreo anidado ‘nested sampling’ y muestreo a intervalo fijo ‘fixed interval sampling’.

Hay otros problemas que influyen en el plan de una muestra espacial. Aumentando la precisión del estimador a partir del aumento en el tamaño de la muestra, siempre sube los costos pero los costos extras de acumular grandes volúmenes de datos pueden ser particularmente importantes en el muestreo espacial. Puede haber problemas de accesibilidad a ciertos sitios y asociados costos de transporte con el muestreo de una población geográficamente dispersa. Es normalmente necesario

hacer un intercambio entre el criterio económico y el estadístico en el diseño de una muestra.

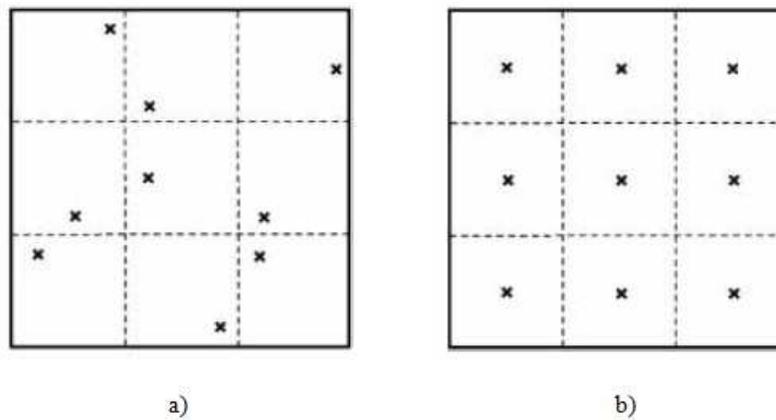


Figura 1.1: Tipos de muestreo espacial.

1.4. Métodos de Interpolación

Los métodos de la interpolación, según Lloyd (2007), pueden ser divididos en dos amplios grupos: los métodos globales y locales. Una distinción ampliamente usada entre los métodos de interpolación es que los métodos globales usan todos los datos para la predicción, mientras los métodos locales normalmente usan sólo algún subconjunto de los datos para hacer cada predicción. Un beneficio de métodos locales es que el tiempo computacional está reducido usando sólo algún subconjunto de los datos para hacer una predicción. Algunos métodos pueden hacer uso de todos los datos disponibles pero pueden tomar en cuenta la distancia de la locación de la predicción. Tales métodos todavía pueden ser considerados locales; tal como, muchas técnicas de interpolación ampliamente usadas son los métodos locales.

Los métodos de la interpolación también pueden ser divididos en dos grupos específicos, interpolación puntual (point interpolation) e interpolación de área (areal interpolation), cada uno con dos subdivisiones.

La interpolación de punto está basada en las muestras disponibles en locaciones específicas que pueden ser aproximados como puntos. Una división extensa comprende métodos exactos y los

métodos aproximados. Los interpoladores exactos hacen una predicción que es igual que el valor real de los datos, sin embargo, los interpoladores aproximados no ajustan los datos de esta manera.

La interpolación de área se usa donde los datos comprenden mediciones realizadas dentro de las áreas y el deseo es, convertir un sistema zonal existente a otro sistema zonal.

Esta sección, se dedica al estudio de la interpolación puntual y se discuten algunas de las técnicas más ampliamente usadas.

1.4.1. Nearest Neighbor (NN)

Un concepto fundamental en la geografía es que entidades cercanas normalmente comparten más similitudes que las entidades que están lejanas. Esta idea se etiqueta a menudo como ‘la primera ley de geografía de Tobler y puede resumirse como *“todo se relaciona a todo lo demás, pero los objetos cercanos están más relacionados que los objetos distantes”* (Miller, 2004). Este hecho implica la autocorrelación espacial para los rasgos en un espacio geográfico, lo que significa que hay una relación entre los rasgos en los puntos vecinos y esto inspira a usar el método del Vecino más Cercano (también conocido como ‘proximal interpolation’ ó ‘point sampling’ en algunos contextos) para la clasificación de puntos de juegos de datos. Este método guarda primero todo el juego de datos y asigna etiquetas (valores). Como consecuencia, asigna el valor del punto más cercano en el juego de datos al nuevo punto. Para determinar qué punto es el punto más cercano para el punto recientemente agregado, el método hace el uso de un diagrama de Voronoi nombrado en honor del matemático ruso Georgy Fedoseevich Voronoi (o Voronoy). Un diagrama de Voronoi es una descomposición de espacio en células (polígonos), uno para cada punto dado, para que en cualquier parte en el espacio, el punto más cercano esté dentro de la célula [ver Figura 1.2].

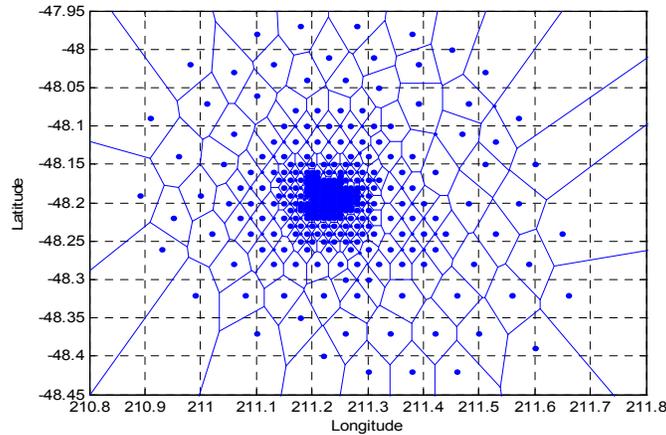


Figura 1.2: Diagrama de Voronoi, tomada de MATLAB R2007b.

El método de vecinos más cercanos puede ser aplicado para diferentes propósitos, por ejemplo, cuando se usa para reconocimiento de patrones en imágenes multi-espectrales se trata de ordenar los píxeles vecinos de acuerdo a una métrica de distancia aplicada a cada uno de los píxeles que componen la imagen multi-espectral (Mathworks, 2009).

A pesar de sus muchas aplicaciones, el método del Vecino más Cercano se critica a menudo porque selecciona el valor del punto más cercano solo, y cada predicción se basa en sólo una medida, no hay ninguna estimación del error, y la información de los puntos vecinos se ignora.

1.4.2. Natural Neighbor Interpolation (NNI)

La Interpolación del Vecino Natural derivada por Robin Sibson, matemático británico, combina los mejores rasgos de los polígonos de Thiessen (los polígonos de Voronoi (Voronoi polygons), Dirichlet tessellation) y Triangulación. El primer paso es una triangulación de los datos por el método de Delaunay en que cada punto en el juego de datos se conecta con una línea recta a todos sus vecinos más cercanos formando triángulos [ver Figura 1.3a]. Los polígonos de Thiessen se forman después circunscribiendo estos triángulos, o sea, se hace una circunferencia circunscrita [ver Figura 1.3b]. Cada polígono de Thiessen se construye conectando todo los puntos centrales (circuncentros) de los triángulos circunscritos [ver Figura 1.3c].

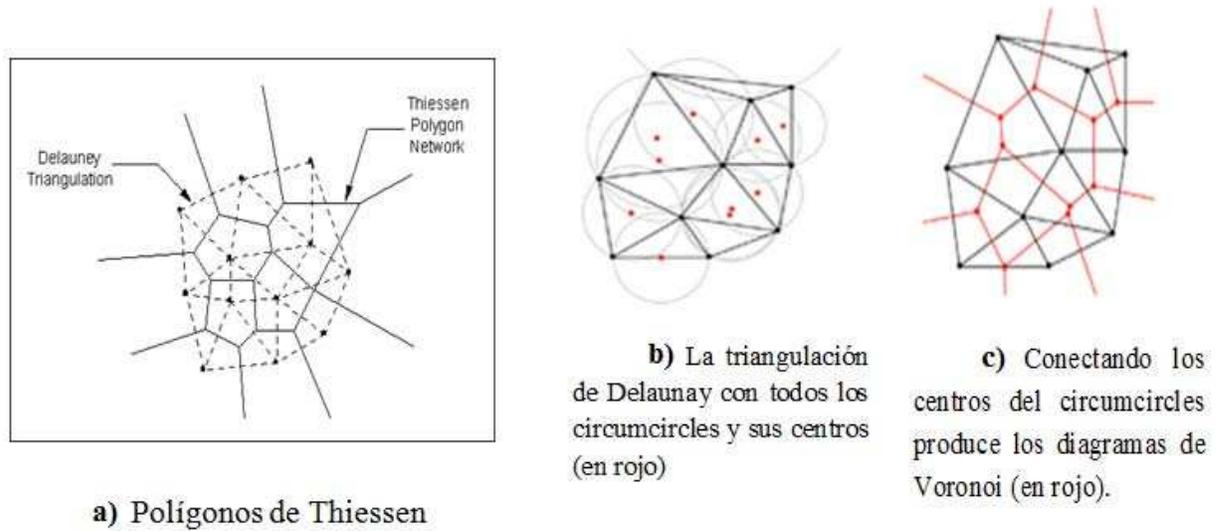


Figura 1.3: Polígonos de Thiessen.

Para determinar el valor de cualquier otro punto (x_0), se toma S un juego de n puntos en el espacio d -dimensional. La célula de Voronoi de un punto $p \in S$, definido como V_p , es el juego de puntos x que es más cercano a p que a cualquier otro punto en S . La unión de las células de Voronoi de todos los puntos generadores p en S forma el diagrama de Voronoi (VD) de S . En ese caso, los vecinos naturales (Natural Neighbors) de x son los puntos en S cuya célula de Voronoi se modificaría si el punto x se inserta en $VD(S)$. La inserción de x crea una nueva célula de Voronoi V_x que “tomo” del volumen de las células de Voronoi que “serían” sus vecinos naturales, como se muestra en la Figura 1.4. Esta idea forma la base de las coordenadas de los vecinos naturales que define cuantitativamente la cantidad V_x que se quita de cada uno de sus vecinos naturales. Sea D el $VD(S)$, y $D^+ = D \cup x$. La célula de Voronoi de un punto p en D se define por V_p , y V_p^+ su célula en D^+ . La coordenada del vecino natural de x con respecto a un punto p_i es dado por la Equación 1.4.1:

$$w_i(x) = \frac{Vol(V_{p_i} \cap V_x^+)}{Vol(V_x^+)} \quad (1.4.1)$$

donde $Vol(V_{p_i})$ representa el volumen de V_{p_i} . Para cualquier x , el valor de $w_i(x)$ siempre estará entre 0 y 1: 0 cuando p_i no es un vecino natural de x , y 1 cuando x está exactamente en la misma posición que p_i . Una consideración importante es que la suma de los volúmenes tomados de cada

uno de los k vecinos naturales es igual a $Vol(V_x^+)$. Por consiguiente, a más alto valor de $w_i(x)$, más fuerte es “la influencia” de p_i sobre x . Las coordenadas de los vecinos naturales se influncian tanto por la distancia de x a p_i como por la distribución espacial de p_i alrededor de x .

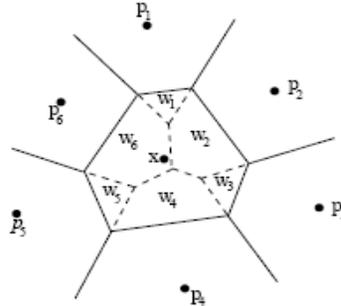


Figura 1.4: Las coordenadas de los vecinos naturales de x .

1.4.3. Inverse distance weighted method (IDW)

El método de la Distancia Inversa Ponderada es un método de interpolación exacto. El objetivo del método es introducir el mínimo de error en la interpolación, para esto utiliza pesos usando distancias que se refieren al punto de partida de la partícula. Cada nueva posición tiene una influencia local que disminuye y que es insignificante después de cierto límite. La influencia atribuida a un dato es radial, simétrica y centrada en un punto. En otras palabras, cuando el valor del dato disminuye con el incremento de la distancia, el peso (λ_i) asignado disminuye atribuyendo menor importancia a los puntos más lejanos. Así el valor de un punto es la suma de esas influencias en cada punto.

Si denotamos por $Z(x_i)$ a la observación en la i -ésima estación, el valor interpolado en el punto (x_0) viene dado por la Ecuación 1.4.2:

$$\bar{Z}(x_0) = \frac{\sum_{i=1}^n Z(x_i) \cdot d_{i0}^{-r}}{\sum_{i=1}^n d_{i0}^{-r}} \tag{1.4.2}$$

donde la predicción hecha en la posición x_0 es una función de las n observaciones vecinas, $Z(x_i)$, $i = 1, 2, \dots, n$; el parámetro r es un exponente que determina el peso asignado a cada una de las observaciones, y d es la distancia en que la posición de la predicción x_0 y la posición de las x_i

observaciones están separadas. Se hace notar que la suma de los pesos es igual a 1 y el valor más comúnmente asignado a r es 2 para que los datos se pesen inversamente al cuadrado de la distancia (García, 2007).

1.4.4. Interpolación basada en Splines

El Spline estima valores usando una función matemática que minimiza la curvatura global de la superficie. Esto produce una superficie lisa que atraviesa de forma exacta todos los puntos. Conceptualmente, es como doblar una hoja de caucho para que atraviese los puntos mientras se minimiza la curvatura total de la superficie (Childs, 2004).

Los Splines usan el método de interpolación local (Piecewise) o por tramos en lugar de la interpolación global. Este método se basa en construir un polinomio de bajo grado entre cada par de puntos de los datos conocidos. Si se usa un polinomio de primer grado, entonces se conoce como interpolación lineal. Cuando los polinomios son de segundo y tercer grado, se llaman splines cuadrático y cúbico respectivamente. Mientras más alto es el grado del spline, es más alisada la curvatura (Blanco, Álvarez, Montenegro, & Gómez, 1998).

1.4.5. Interpolación basada en Kriging

El Kriging, nombrado así en honor de un minero sudafricano, el ingeniero D.G. Krige en los años cincuenta, es un método geoestadístico muy difundido y del cual existen un gran número de modificaciones. El método está basado en la descripción de la variación espacial de los datos que se modelan mediante el variograma. Este se calcula por la expresión 1.4.3 (García, 2007):

$$2\gamma^*(h) = \frac{1}{n} \sum [g(x) - g(x+h)]^2 \quad (1.4.3)$$

donde h es la distancia entre los puntos y $\sum [g(x) - g(x+h)]$ es la diferencia en el valor de todas las muestras separadas por una distancia h .

Los modelos utilizados habitualmente en el ajuste del variograma son: Exponencial, Lineal y Esférico (Clarke, 2001).

▷ Esférico:

$$\begin{aligned} \gamma(h) &= C \left(\frac{3h}{2a} - \frac{1h^3}{2a^3} \right) \text{ donde } h \leq a \\ &= C \text{ donde } h \geq a \end{aligned}$$

La distancia a la que las muestras se ponen independientes (la autocorrelación desaparece o casi desaparece) entre sí, se denota por a y se llama rango de influencia (range of influence) de una muestra. El valor donde el gráfico se nivela se denota por C y se llama “sill” del semi-variograma (ver la Figura 1.5).

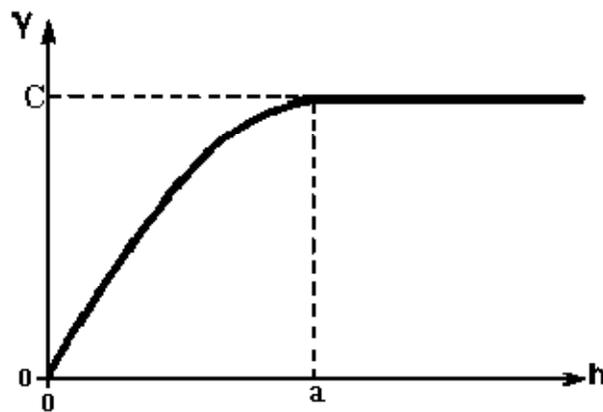


Figura 1.5: La forma “ideal” para un semi-variograma: modelo esférico.

▷ Exponencial:

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

▷ Lineal:

$$\gamma(h) = ph, \text{ donde } p \text{ es la pendiente}$$

▷ Existe otro modelo, para describir el semi-variograma de un fenómeno completamente aleatorio. Eficazmente, es un modelo esférico con un rango muy pequeño de influencia. El

“nugget effect” así expresado se refiere a la situación en que la diferencia entre las medidas tomadas en locaciones de muestras muy cercanas, no es cero (Krivoruchko & Gribov, 2000), y está dado por:

$$\gamma(h) = 0$$

$$\gamma(h) = C_0 \text{ donde } h > 0$$

Con la adición del “nugget effect”, todos los modelos del semi-variograma necesitarán ser modificados a la forma:

$$\gamma(h) = 0$$

$$\gamma(h) = C_0 + \gamma(h) \text{ donde } h > 0$$

El valor interpolado en un punto x_0 viene dado por la combinación lineal:

$$\bar{Z}(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) \quad (1.4.4)$$

donde (λ_i) es el peso asignado a cada muestra $Z(x_i)$. En el Kriging los pesos son escogidos para que el estimador $\bar{Z}(x_0)$ del valor verdadero $Z(x_0)$ sea no sesgado y la predicción $\sigma^2(x_0)$ sea mínima (Kalivas, Triantakostas, & Kollias, 2002). Eso es:

$$E \left[\bar{Z}(x_0) - Z(x_0) \right] = 0 \quad (1.4.5)$$

$$\sigma^2(x_0) = \text{var} \left[\bar{Z}(x_0) - Z(x_0) \right] = \text{mínimo} \quad (1.4.6)$$

Para asegurar que la predicción sea insesgada los pesos asignados a cada punto vecino deben satisfacer:

$$\sum_{i=1}^n \lambda_i = 1 \quad (1.4.7)$$

Hay numerosos métodos y aproximaciones a la interpolación. Cada método usa una técnica diferente para determinar el valor aproximado de los puntos de datos. El método más apropiado dependerá de la distribución de puntos de la muestra y del fenómeno a ser estudiado. Aquellos

que se mencionan debajo forman un subgrupo importante y ampliamente utilizados en la ciencia medioambiental. Ellos no constituyen el centro de esta investigación, pero una mención rápida es digna. Los métodos más notables son: Spatial Moving Average (SMA), Triangulated Irregular Network (TIN), Angular Distance Weighting (ADW), Angular Distance Weighting (ADW), Cressman, Trend Surfaces, Topo to Raster, la Interpolación Condicional, Reduced Space Optimal Interpolation, la Regresión Lineal y Polinómica.

1.5. Métodos de Validación

Antes de tomar cualquier decisión en base a los métodos de la interpolación, es crítico evaluar qué bien ellos la ejecutan. En cuanto a la interpolación espacial, un acercamiento razonable es comparar las predicciones geoestadísticas con observaciones que se han quitado temporalmente una tras la otra (Cross-validation) o extraer para el análisis entero (Jackknife) (Goovaerts, 2006) o hacer sucesivas muestras de n juegos de datos de la población N de datos (Bootstrap).

Aunque existen varios métodos para calcular la validez de la correlación de Pearson, según Trauth (2006), los métodos de hacer muestreos sucesivos “*re-sampling schemes*” como el Bootstrap y Jackknife, proveen una vía alternativa para obtener la validez estadística del resultado de aplicar un método de interpolación de datos. El método de Bootstrap toma un subconjunto de muestras de forma repetida de los datos originales. La estadística de estos subconjuntos de muestras provee mejor información de las características de la población que los parámetros estadísticos tradicionales, (media, desviación estándar, coeficientes de correlación) aplicada sobre la población entera.

El proceso Cross-validation (“*leave-one-out*”) puede resumirse más allá en los pasos siguientes:

- i. Un punto de dato es quitado del juego de datos, uno tras el otro.
- ii. Su valor se predice usando los datos restantes (o los datos vecinos).
- iii. El valor real para el punto se compara con el valor predicho.

iv. El punto de datos es retornado al juego de datos.

Para el “Cross-validation”, es necesario considerar la correspondiente salida cuando la i -ésima muestra se omite. Un pseudo-valor Z_i^+ correspondiente a cada punto omitido, se calcula como sigue:

$$Z_i^+ = \sum_{i=1}^n \left[z(x_i) - \bar{Z}(x_{\neq i}) \right] \quad (1.5.1)$$

El parámetro $\bar{Z}(x_{\neq i})$ es el valor de interpolación de la variable en el punto i , estimado de los $n - 1$ puntos restantes, $z(x_i)$ es el valor actual medida de la variable quitada desde el punto i , n es el número de puntos.

Según Lloyd (2007), este proceso continúa hasta que todos los puntos de datos se hayan predicho de esta manera y pueden obtenerse las estadísticas de error de predicción. Una de las formas de comparación es el promedio de la raíz cuadrada de los errores de predicción, que es el promedio de las diferencias cuadradas entre los valores reales y los valores predichos.

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N \left[z(x_i) - \bar{Z}(x_{\neq i}) \right]^2} \quad (1.5.2)$$

En el Jackknife (Tomczak, 1998), todas las n muestras (locaciones medidas) se usan para estimar el parámetro Z al nodo interpolado “ j ” (Z_{ALL}). Luego el Jackknife procede quitando una observación tras la otra del juego de datos original, y estimando el valor de parámetro Z repetidamente al nodo “ j ” de los $(i - 1)$ puntos que permanecen en los datos. El Jackknife es calculado utilizando la ecuación 1.5.3.

$$Z_i^+ = n \cdot Z_{ALL} - (n - 1) \cdot Z_{-i} \quad i = \overline{1, n} \quad (1.5.3)$$

donde n es el tamaño de la muestra, Z_{ALL} es el parámetro de estimación para el nodo j usando todos los n puntos de datos, Z_{-i} es el parámetro de estimación cuando la i -ésima muestra es quitada, Z_i^+ es una estimación del pseudo-valor para el nodo “ j ” que corresponden a los i -ésimos datos apuntan al ser alejado.

El estimador Jackknife del parámetro Z a la posición “ j ” es la media de todo los pseudo-valores para el nodo “ j ”:

$$Z_j = \frac{\sum_{i=1}^n Z_i^+}{n} \quad (1.5.4)$$

donde (Z_j) es el estimador Jackknife del parámetro Z .

El procedimiento del Jackknife se repite y los valores de Z_j y σ_J son calculados para cada una de las posiciones estimadas del nodo “ j ”. El uso de pseudo-valores permite evaluar la precisión del estimador Jackknife a través del error estándar normal estimado σ_J que se define como:

$$\sigma_J = \sqrt{\left(\frac{1}{n \cdot (n-1)}\right) \cdot \sum_{i=1}^n (Z_i^+ - Z_j)^2} = \sqrt{\left(\frac{(n-1)}{n}\right) \cdot \left(\sum_{i=1}^n (Z_i^+)^2 - \frac{1}{n} \left(\sum_{i=1}^n Z_i^+\right)^2\right)} \quad (1.5.5)$$

1.5.1. Comparación de los métodos de interpolación

Como se declaró antes, el propósito de esta investigación es comparar varios métodos de interpolación usados en la ciencia medioambiental para determinar cuál de ellos ofrece la mejor aproximación para los datos espaciales. Con el uso de técnicas de comprobación (Validation) se obtuvieron los dos métodos más eficaces, que serán como consecuencia el foco del resto de esta investigación.

En un estudio dirigido en la comparación de los métodos de la interpolación para el mapeo de variables de temperatura (Kurtzman & Kadmon, 1999), los resultados sugieren que en algunos casos los modelos de la regresión globales simples puedan ser tan eficaces como los métodos de interpolación locales sofisticados, sobre todo al tratar con variables ambientales. Sin embargo, para encontrar el mejor estimador (interpolador) en general, el resumen de los resultados de varios estudios se presenta como sigue.

En una comparación entre el Vecino más Cercano (Nearest Neighbor), Inverse Distance Weighting, Regresión Lineal e Interpolación Polinómica (Pecušová, Parajka, & Hrušková, 2002), los resultados indican que el método de regresión polinómica tiene la habilidad de predecir de

forma más fiable entre los esquemas de la interpolación seleccionados. Sin embargo, al comparar la interpolación de la Regresión Polinómica con la interpolación de Spline (Breaz, 2004), los resultados indican que la interpolación basada en los Splines produce la mejor aproximación. En estudios adicionales dirigidos (Chegini, Hosseini, Mahdian, & Hossein, 2001) (Mahdian, Bandarabady, Sokouti, & Banis, 2009), entre el Spline y otros métodos de la interpolación, los resultados sugieren que la interpolación de Splines ofrece estimaciones con mejor precisión.

En una investigación titulada “ *The selection of the most appropriate interpolation method for sandstone reservoirs in the Kloštar oil and gas field in Croatia* ” (La selección del método de la interpolación más apropiado para los depósitos de la piedra arenisca en el campo de gas y petróleo de Kloštar en Croacia) (Balic, Velic, & Malvic, 2008) los resultados indicaron que la interpolación basada en Kriging proporciona la mejor estimación cuando se comparó con los métodos de interpolación Inverse Distance Weighting y Nearest Neighbor. Los resultados también muestran que el Inverse Distance Weighting es un mejor interpolador que el Nearest Neighbor.

Una comparación de métodos de interpolación espaciales para la estimación de datos de calidad del aire fue hecha (Wong, Yuan, & Perlin, 2004) para Kriging, Spatial Averaging, Nearest Neighbor e Inverse Distance Weighting. Los resultados sugieren que la interpolación de Kriging proporciona la mejor estimación.

La mayoría de la literatura consultada indica que Kriging (Fencík & Vajsáblová, 2006), (Hofstra, Haylock, Jones, & Frei, 2008), (Wong *et al.*, 2004) o Splines (Mahdian *et al.*, 2009), (Chegini *et al.*, 2001), proponen la mejor estimación de los datos medioambientales espaciales.

1.6. Conclusiones parciales

En el presente capítulo se validó la importancia del tema haciendo un estudio del estado actual de interpolación de datos espaciales ambientales, los métodos de muestreos y la importancia de realizar tales interpolaciones. Una comparación de varios métodos de la interpolación usados en el análisis de los datos espaciales medioambientales ha resultado en la selección de la interpolación basada en el Kriging y los Splines como los dos métodos que proporcionan la mejor aproximación.

Capítulo 2

Materiales y Métodos

2.1. Interpolación basada en Splines

2.1.1. Teoría

El ajuste de una curva polinómica a un juego de puntos de datos tiene aplicaciones en el CAD (Diseño Asistido por Computadoras), CAM (la Manufactura Asistida por Computadora), y sistemas de gráficos de computadora. Un operador quiere dibujar una curva lisa a través de puntos de datos que no están sujetos al error. Tradicionalmente, era común unir los puntos y subjetivamente dibujar una curva que pareciera lisa ante la vista (Mathews & Fink, 1999). Matemáticamente, podría parecer que un polinomio interpolador sería la solución idónea, pero el hecho de que ese polinomio tendrá grado n , significa una seria dificultad: lograr que la curva pase por un gran número de puntos sería muy trabajoso de operar e incluso evaluar.

En casos como este, es preferible abordar el problema de otra manera: en lugar de buscar un solo polinomio que satisfaga las $n + 1$ condiciones:

$$P(x_i) = y_i \quad i = \overline{0, n}$$

Hallar varios polinomios de menor grado, cada uno de cuales pase por un número reducido de nodos (puntos) consecutivos o sea se hace una interpolación por tramos (interpolación local). En el caso de splines cúbicos (Kiusalaas, 2005), se hace un polinomio de grado 3 en cada sub-intervalo (tramo) de nodos o sea una función cúbica $S_i(x) = f_{i, i+1}(x)$ en cada subintervalo $[x_i, x_{i+1}]$, ver la Figura 2.1.

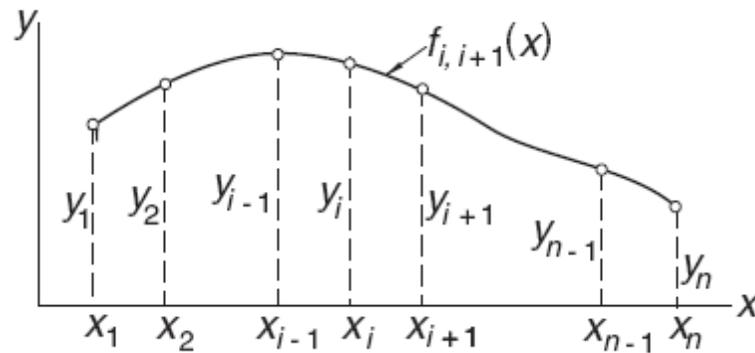


Figura 2.1: Splines Cúbicos.

2.1.2. Formulación Matemática: Splines Cúbicos

Las Propiedades de Splines Cúbicos

Suponga que $x_i, y_i, (i = \overline{0, n})$ son $N + 1$ puntos, donde $x_0 < x_1 < \dots < x_N$. Un spline cúbico $S(x)$ se define por las siguientes condiciones:

I. Donde cada $S_i(x)$ es un polinomio cúbico, o sea:

$$S_i(x) = a_i x^3 + b_i x^2 + c_i x + d_i$$

II. Cuando se utiliza $S(x)$ como interpoladora debe tomar en los nodos de interpolación idénticos valores (interpolación exacta), o sea:

$$S(x_i) = y_i \quad \text{para } i = \overline{0, N}$$

III. $S(x)$ es continua en $[x_0, x_n]$:

$$S_i(x_{i+1}) = S_{i+1}(x_{i+1}) \quad \text{para } i \in [0, N - 2]$$

IV. $S'(x)$ es continua en $[x_0, x_n]$.

V. $S''(x)$ es continua en $[x_0, x_n]$.

En términos geométricos, la gráfica de $S(x)$ está formada por n segmentos de polinomios cúbicos de tal manera que la curva es continua, la pendiente varía en forma continua (no hay puntos con ángulos pronunciados) y la curvatura varía en forma continua.

Para construir un spline cúbico que satisfaga las condiciones I hasta V , hay que tener en cuenta:

Cantidad de incógnitas:

Cada S_i tiene 4 incógnitas (a_i, b_i, c_i, d_i) . Entonces, hay un total de $4n$ incógnitas.

Número de condiciones:

El spline tiene que interpolar todos los $n + 1$ puntos:

$$\triangleright S_i(x_i) = y_i \quad \text{para } 0 \leq i \leq n - 1 ; S_{n-1}(x_n) = y_n \quad \text{da } (n + 1) \text{ condiciones}$$

Las condiciones de continuidad para el spline, su derivada, y su segunda derivada son las siguientes:

$$\triangleright S_{i-1}(x_i) = S_i(x_i) \quad \text{para } 1 \leq i \leq n - 1 \text{ da } (n + 1) \text{ condiciones ;}$$

$$\triangleright S'_{i-1}(x_i) = S'_i(x_i) \quad \text{para } 1 \leq i \leq n - 1 \text{ da } (n + 1) \text{ condiciones ;}$$

$$\triangleright S''_{i-1}(x_i) = S''_i(x_i) \quad \text{para } 1 \leq i \leq n - 1 \text{ da } (n + 1) \text{ condiciones ;}$$

El resultado de sumar todas estas condiciones es $4n - 2$ condiciones.

Para obtener una solución única o sea $4n$ incógnitas con $4n$ ecuaciones, se debe imponer 2 condiciones más. Una vía común es hacer lo siguiente:

$$S''_{i-1}(x_0) = S''_i(x_n) = 0$$

La asignación de las dos derivadas a cero resulta en una función llamada *Splines Cúbicos Naturales*.

2.1.3. El cálculo de un Spline Cúbico Natural

Para determinar los $4n$ coeficientes de la función $S(x)$, se seguirá el procedimiento de ir imponiendo sucesivamente las $4n$ condiciones ya señaladas. Para facilitar los cálculos, se usará la siguiente notación:

$$z_i = S''(x_i) \quad \text{para } 0 \leq i \leq n$$

$$h_i = x_{i+1} - x_i \quad i = 0, 1, 2, \dots, n - 1$$

Como se mencionó anteriormente una de las condiciones era que las segundas derivadas sean iguales, o sea $S''(x_0) = S''(x_n) = 0$, se puede decir que $z_0 = z_n = 0$. Sabiendo esto, lo que queda es buscar las demás z_i en el intervalo $1 \leq i \leq n - 1$ haciendo lo siguiente:

Como en cada subintervalo $[x_i, x_{i+1}]$, $S_i(x)$ es un spline cúbico de la forma:

$$S_i(x) = a_i x^3 + b_i x^2 + c_i x + d_i$$

Su segunda derivada será en ese intervalo, una función lineal, esto es, para $x_i \leq x \leq x_{i+1}$

$$S''_i(x) = z_i + m_i(x - x_i) \tag{2.1.1}$$

donde $m_i = \frac{z_{i+1} - z_i}{h_i}$, es la pendiente de la recta.

Simplificando la ecuación 2.1.1 resulta en lo siguiente:

$$S''_i(x) = \frac{z_{i+1}}{h_i}(x - x_i) + \frac{z_i}{h_i}(x_{i+1} - x) \tag{2.1.2}$$

Sabiendo el valor de z_i , se puede calcular fácilmente la segunda derivada $S''_i(x)$, del spline cúbico, pero el interés es en el spline $S(x)$ no en la segunda derivada.

Para calcular $S(x)$, se integra dos veces la ecuación 2.1.2 dando:

$$S_i(x) = \frac{z_{i+1}}{6h_i}(x - x_i)^3 + \frac{z_i}{6h_i}(x_{i+1} - x)^3 + C_i(x - x_i) + D_i(x_{i+1} - x) \tag{2.1.3}$$

para $x_i \leq x \leq x_{i+1}$

donde C_i y D_i son constantes de la integración.

Para resolver C_i y D_i , se utiliza la ecuación 2.1.3 con las condiciones $S_i(x_i) = y_i$ y $S_i(x_{i+1}) = y_{i+1}$. Para una deducción de la ecuación 2.1.3 y las siguientes ecuaciones C_i y D_i , ver páginas 264 - 265 (Blanco et al., 1998). Esto da como resultado de un sistema lineal de ecuaciones:

$$C_i = \frac{y_{i+1}}{h_i} - \frac{h_i}{6} z_{i+1} \quad (2.1.4)$$

$$D_i = \frac{y_i}{h_i} - \frac{h_i}{6} z_i \quad (2.1.5)$$

Sustituyendo ecuación 2.1.4 y 2.1.5 en la ecuación 2.1.3 y agrupando términos, se llega a:

$$\begin{aligned} S_i(x) &= \frac{1}{6h_i} [z_i(x_{i+1} - x)^3 + z_{i+1}(x - x_i)^3] \\ &+ \frac{1}{h_i} [(x_{i+1} - x)y_i + (x - x_i)y_{i+1}] \\ &- \frac{h_i}{6} [(x_{i+1} - x)z_i + (x - x_i)z_{i+1}] \end{aligned} \quad (2.1.6)$$

Se puede calcular el spline cúbico con la ecuación 2.1.6, si y solo si se tiene z_i .

El cálculo de z_i

Para calcular z_i para $1 \leq i \leq n - 1$, se necesita utilizar la última condición de continuidad, o sea $S'_{i+1}(x_i) = S'_i(x_i)$, calculando la derivada de $S'_i(x)$ en la ecuación 2.1.3 dando:

$$S'_{i-1} = \frac{h_{i-1}}{6} z_{i-1} + \frac{h_{i-1}}{3} z_i + b_{i-1} \quad (2.1.7)$$

donde $b_i \equiv \frac{1}{h_i}(y_{i+1} - y_i)$

Esta condición de la continuidad de la primera derivada en forma simplificada (ver páginas 265-266 (Blanco et al., 1998) para la deducción de la fórmula), da como resultado:

$$h_{i-1}z_{i-1} + 2(h_{i-1} + h_i)z_i + h_i z_{i+1} = 6(b_i - b_{i-1}) \quad (2.1.8)$$

2.2. Interpolación basada en Kriging

La mayoría de las propiedades del ambiente podrían ser medidas en cualquiera de un número infinito de lugares, pero en la práctica solo se puede tomar un número limitado de observaciones, principalmente por razones de costos. Si se desea saber los valores en otras partes, entonces se debe estimar de los datos observados, de igual manera sucede si se quiere estimar en áreas más grandes donde no ha sido posible medir directamente. Típicamente estos lugares son los nodos de una malla regular (regular grid) puestos en el dominio estudiado, el proceso de interpolación en este caso es conocido como el “gridding.” Una vez que se establecen las mallas, se usa a menudo como la representación de la realidad, sin referencia a los datos originales. Pero, los mayores esfuerzos en la predicción espacial han sido en forma matemática, basados en la geometría y alguna apreciación de la naturaleza física de los fenómenos. La mayoría toma en cuenta sólo la variación sistemática o determinista, pero ninguno toma en cuenta el error. La predicción geoestadística basada en Kriging es la conclusión lógica de estos esfuerzos en el sentido de que se construye sobre ellos y supera sus debilidades.

2.2.1. Teoría

La geoestadística es una rama de la estadística aplicada que se especializa en el análisis, modelación y predicción de la variabilidad espacial de fenómenos en Ciencias de la Tierra (Chiles & Delfiner, 1999). La geoestadística fue desarrollada originalmente para la estimación, y generalmente se llama *Kriging* en honor a D.G. Krige, ingeniero minero de Suráfrica, aunque el término se acuñó originalmente como *krigeage* por P. Carlier, pero Matheron en 1963 lo dio a conocer en idioma inglés en reconocimiento a la contribución de Krige a mejorar la precisión de estimar la concentración de oro y otros metales en la minería. Aunque mucho del crédito por formalizar la técnica va a Matheron y sus colegas, la matemática de Kriging simple fue deducida por A. N. Kolmogorov en los años treinta (Webster, 2007).

Según Trauth (2006), un supuesto básico en la geoestadística es que un proceso que ocurre en el

espacio y tiempo (espacio-temporal) está compuesto por componentes deterministas y estocásticos. Los componentes deterministas pueden ser las tendencias globales y locales (drifts). El componente estocástico se forma por una parte completamente al azar y una parte de autocorrelación. Un componente de autocorrelación implica que como promedio, las observaciones más cercanas son más similares que las observaciones más distantes. Esta conducta se describe por medio del variograma donde las diferencias cuadráticas entre las observaciones se trazan contra su distancia de separación. La idea fundamental de D. Krige era usar el variograma para la interpolación como medio para determinar la magnitud de influencia de las observaciones vecinas al predecir los valores de las locaciones (puntos) desconocidas. La geoestadística lineal básica incluye dos procedimientos principales: “variography” para modelar el variograma y “Kriging” para la interpolación.

Variography

Variography¹ según Clarke (2001), relacionado con el variograma mencionado en el epígrafe 1.4 puede ser explicado como sigue: “...si se considera el problema de estimación local, por ejemplo, se intenta estimar el valor a asociado al punto A en la Figura 2.2, dadas varias muestras en una vecindad. Es razonable utilizar un procedimiento de estimación que de más importancia a la muestra 1 (más cercana) que a la muestra 5 (más alejada). Existe una amplia variedad de métodos para decidir “el peso” otorgado a cada muestra, principalmente basado en la distancia entre la muestra y el punto a ser estimado. Los valores de la muestra pueden ser ponderados por el inverso de la distancia, el inverso de la distancia cuadrática, o por alguna constante arbitraria (e.j: el rango de influencia) menos la distancia. Todos estos casos involucran la misma suposición básica, o sea, la relación entre el valor al punto A y cualquier valor de la muestra dependiendo de la distancia (y posiblemente la dirección) entre las dos posiciones, y en nada más ...”.

¹La técnica de modelar el variograma



Figura 2.2: El muestreo hipotético y situación de estimación.

En la Figura 2.2, parece lógico esperar que el valor en la posición 5 sea “muy diferente” del valor en la posición de estimación A , mientras la muestra 1, tendrá un valor “no muy diferente” del valor en A . Se puede suponer que la diferencia en el valor entre dos posiciones en el depósito sólo depende de la distancia entre ellos y sus orientaciones relativas. Por ejemplo, si se toma un par de muestras separadas por 50 pies en una línea norte-sur en una parte del depósito, y se mide la diferencia entre los valores. Ahora, supóngase que se hace lo mismo, 200 pies más lejos y nuevamente en otra posición a una distancia de separación h , y así sucesivamente.

Por simplicidad, se denota la distancia entre las muestras como h (ver la Figura 2.3).

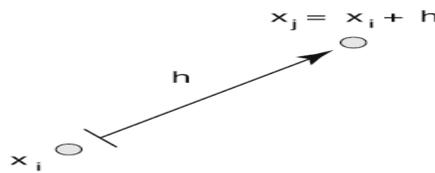


Figura 2.3: Vector de separación h entre dos puntos.

Como se ha dicho, la diferencia en el valor entre las dos muestras sólo depende de h , en estas condiciones estadísticas, la distribución de las diferencias depende sólo de h . Si este planteamiento se cumple para la distribución, también sucede para la media y varianza. Es decir, se puede describir la diferencia media en el valor de la muestra como $m(h)$ y la varianza de estas diferencias como $2(h)$. Así, con un juego de pares de muestras para un h específico (ej. 50 pies, norte-sur) se puede

calcular un valor “experimental” para $m(h)$.

$$m^*(h) = \frac{1}{n} \sum [g(x) - g(x + h)] \quad (2.2.1)$$

donde g es el valor de la muestra dada en la posición x , x denota la posición del primer punto de la muestra en un par $[x; x + h]$ y h la distancia del siguiente punto, y n es el número total de pares de muestras. La introducción del “*” en una variable es para mostrar que ha sido calculada en lugar de definido “teóricamente”. Un mirada más de cerca a $m(h)$ muestra que se representa una media de la diferencia de los valores entre dos muestras, en otros términos, una diferencia “esperada”. Si $m(h)$ es cero, implica que no se “espera” ninguna diferencia entre los valores de muestras separadas por una distancia h . Dicho de otra manera, localmente (dentro de h) no hay ninguna tendencia.

Volviendo a la variación de las diferencias, que quedó definido como $2\gamma(h)$ y normalmente se conoce como variograma, porque varía con la distancia (y dirección) h .

$$2\gamma^*(h) = \frac{1}{n} \sum [g(x) - g(x + h)]^2 \quad (2.2.2)$$

El número “2” delante de γ está allí por conveniencia matemática, $\gamma(h)$ representa el semi-variograma y $\gamma^*(h)$ es el semi-variograma experimental.

Kriging

Kriging proporciona una solución al problema de estimación basado en un modelo continuo de variación espacial estocástica. Este método hace el mejor uso del conocimiento existente sobre la vecindad al punto de interés tomando cuenta la manera que una propiedad varía en el espacio a través del modelo del variograma. En su formulación original una estimación del Kriging a un lugar era simplemente una suma lineal o el promedio ponderado de los datos en su vecindad.

El Kriging se ha elaborado para resolver problemas complejos en la minería, la ingeniería petrolera, el control y disminución de polución, y la salud pública. El término se utiliza de forma genérica, pero se cuenta con varios tipos distintos de Kriging lineal y no lineal. En esta tesis, el énfasis está en los métodos lineales más simples.

En el Kriging lineal las estimaciones son combinaciones lineales de pesos asignados a los datos.

Los pesos se asignan a los datos de la muestra dentro de la vecindad del punto o bloque estimado de forma tal que se minimice el error de estimación o variación del Kriging, y las estimaciones son imparciales. El Kriging es óptimo en este sentido (Webster, 2007).

2.2.2. Formulación Matemática

Continuando la idea introducida en el epígrafe 1.4 sobre el Modelo de Variograma y Kriging, se presenta la base matemática para el proceso de Kriging. Con la ayuda de Kriging ordinario, el tipo de Kriging más practicado en la actualidad, se presenta la formulación matemática del proceso de Kriging.

Si se retoma la explicación dada anteriormente sobre el término **variography** a partir de la existencia de un punto o área del que no se sabe su valor, pero se desea estimar.

Sea este valor “desconocido” representado por T , y el área o punto de interés representado por A . Según Clarke (2001), para producir un estimador, se debe tener un poco de información, normalmente en forma de muestras. Se supone que existen n muestras con los valores de $g_1, g_2, g_3, \dots, g_n$. Este juego de muestras generalmente es denotado por S y de estas se puede formar un tipo de estimador “lineal”, es decir, un promedio de pesos.

El estimador se denota por T^* y es igual a:

$$T^* = w_1g_1 + w_2g_2 + w_3g_3 + \dots + w_n g_n \quad (2.2.3)$$

Esta misma idea dada en notación compacta sería:

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

donde $Z(x_0) = T^*$; $z(x_0) = g_i$ y $\lambda_i = w_i$

Para que el estimador sea insesgado se impone la condición:

$$\sum_{i=1}^n \lambda_i = 1$$

donde los $w_1, w_2, w_3, \dots, w_n$ son los pesos asignados a cada muestra. Los métodos de estimación locales en la actualidad usan técnicas de medias ponderadas como la distancia inversa donde los pesos suman 1.

Retomando la Figura 2.2 utilizada en la explicación de “Variography” en el epígrafe 2.2.1, se muestra el punto de interés que queda en la posición A , y los cinco puntos de muestras que quedan en su vecindad. Utilizando el posible procedimiento de estimación más simple. Se toma el valor a la posición de la muestra más cercana (1) y se “extiende” a este punto desconocido. Siguiendo esta idea se produce un error de estimación, ε , que será igual a la diferencia entre el valor real T y el valor estimado T^* , que en este caso es igual a g_1 . Eso es:

$$T^* = g_1$$

$$\varepsilon = T - T^*$$

No es demasiado difícil mostrar que si no hay ninguna tendencia (por lo menos localmente), en otros términos, si los valores no son deterministas y allí existe una autocorrelación entre las muestras, este estimador es imparcial. Es decir, si se hacen muchas estimaciones similares, el error tiende a ser cero. Este tipo de estimador es conocido como estimador “lineal” porque es una combinación lineal de los valores de la muestra.

$$\bar{\varepsilon} = 0$$

$$E[Z(x_0) - Z(x_0)] = 0$$

La “fiabilidad” de la estimación puede medirse a partir de la dispersión de los errores. Si los errores toman valores de forma consistente cerca del cero, entonces el estimador es un “bueno”. Si la dispersión de los valores es grande, entonces el estimador no será confiable. La medida estable más simple de la dispersión (estadísticamente) es la desviación normal. La desviación normal de un error de estimación - o el error estándar, es comúnmente utilizada para medir la fiabilidad de ese estimador. La variación de estimación para el estimador “insesgado lineal” general es:

$$\begin{aligned} \text{var}[Z(x_0)] &= E[Z(x_0) - Z(x_0)]^2 \\ &= 2 \sum_{i=1}^N \lambda_i \gamma(x_i, x_0) - \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j \gamma(x_i, x_j) \end{aligned} \tag{2.2.4}$$

donde $\gamma(x_i, x_j)$ es la semi-varianza de Z entre los puntos de datos x_i y x_j , y $\gamma(x_i, x_0)$ es la semi-varianza entre el i -ésimo punto de dato y el punto estimado x_0 .

Hasta ahora se ha considerado sólo el funcionamiento de promediar todas las muestras locales y aplicar este valor como la estimación del área bajo consideración. En un acercamiento mucho más común y probablemente más realista a la estimación de los valores locales, los investigadores estarían interesados en el valor medio de una región o bloque o algún área A ver la figura 2.4. Existen casos en que no parecería sensible pesar todas las muestras igualmente, porque algunas estarán a una gran distancia del área “desconocida” A .

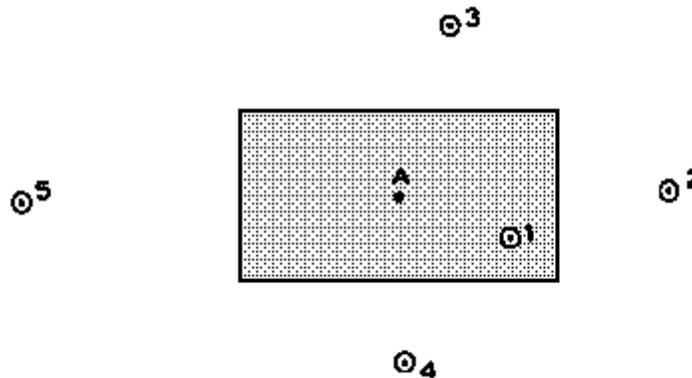


Figura 2.4: Una estimación más realista - el valor del bloque es requerido.

La variación de la estimación todavía es un variograma, pero está ahora el variograma entre los valores en el punto de la muestra (1) y la media de los valores del (panel) área A . La última versión general de la variación de estimación se convierte ahora en:

$$\sigma_\epsilon^2 = 2 \sum_{i=1}^n w_i \bar{\gamma}(S_i, A) - \sum_{i=1}^n \sum_{j=1}^n w_i w_j \bar{\gamma}(S_i, S_j) - \bar{\gamma}(A, A) \tag{2.2.5}$$

donde $\gamma(S, A)$ es el valor promedio del semi-variograma entre cada muestra y el área desconocida, ahora se forma un promedio pesado para cada muestra individual con el área A . El término $\bar{\gamma}(A, A)$ es el semi-variograma promedio entre cada punto en el rectángulo (panel) del área A y cada punto en el rectángulo. Sin embargo, ahora todavía existe otra fuente de variación espuria. El estimador sólo fue considerado como la media de los valores de las muestras, pero $\bar{\gamma}(S, A)$ tiene en cuenta los valores individuales. Así también se tiene que sustraer un término (S, S) de la variación donde éste es el valor promedio del semi-variograma entre cada punto en el juego de muestras y cada punto en el juego de muestras.

Si un estimador “lineal” se usa (un promedio de los pesos de los valores de la muestra), como se declaró previamente, es un estimador insesgado si los pesos suman 1. Hay un número infinito de tales estimadores insesgados lineales, por eso, es necesario buscar “el mejor” de todos, y “el mejor” puede definirse como el que “tiene la más pequeña variación de estimación”. La expresión para la variación de estimación depende de tres cosas: la geometría básica de las muestras y el área desconocida, el tipo de modelo de semi-variograma, y el peso asignado a cada muestra. Dado cualquier procedimiento, la variación solo puede ser cambiada alterando los valores de los pesos. Así, la varianza de estimación debe ser minimizada con respecto a los pesos. La varianza es una función simple de los pesos, esta función debe ser diferenciada y el diferencial igualado a cero para lograr minimizarlo:

$$\frac{\partial \sigma_{\epsilon}^2}{\partial w_i} = 0 \quad i = \overline{1, n} \quad (2.2.6)$$

Esto proporcionará n ecuaciones con n incógnitas $(w_1, w_2, w_3, w_4, \dots, w_n)$. Estos pesos proporcionarán un estimador que tiene el valor mínimo de varianza de estimación. Sin embargo, ellos necesariamente no sumarán uno. No hay nada en el sistema anterior de ecuaciones que limiten los pesos de esta manera, efectivamente la ecuación $\sum w_i = 1$ también necesitará ser satisfecha. Así, para obtener el mejor estimador lineal imparcial (BLUE) - “Best Linear Unbiased Estimator” por sus siglas en inglés, primero hay que satisfacer $(n + 1)$ ecuaciones. Sin embargo, hasta ahora hay n incógnitas que no es una condición muy deseable. Para rectificar esto, otra incógnita debe ser introducida en la forma de multiplicadores de Lagrange, para equilibrar el sistema.

Después de que la diferenciación se ha realizado y las ecuaciones se ordenaron, el sistema resulta en el siguiente sistema de ecuaciones:

$$\begin{array}{cccccc}
 w_1\bar{\gamma}(S_1, S_1)+ & w_2\bar{\gamma}(S_1, S_2)+ & w_3\bar{\gamma}(S_1, S_3)+ & \dots + & w_n\bar{\gamma}(S_1, S_n)+ & \lambda = \bar{\gamma}(S_1, A) \\
 w_1\bar{\gamma}(S_2, S_1)+ & w_2\bar{\gamma}(S_2, S_2)+ & w_3\bar{\gamma}(S_2, S_3)+ & \dots + & w_n\bar{\gamma}(S_2, S_n)+ & \lambda = \bar{\gamma}(S_2, A) \\
 w_1\bar{\gamma}(S_3, S_1)+ & w_2\bar{\gamma}(S_3, S_2)+ & w_3\bar{\gamma}(S_3, S_3)+ & \dots + & w_n\bar{\gamma}(S_3, S_n)+ & \lambda = \bar{\gamma}(S_3, A) \\
 \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
 w_1\bar{\gamma}(S_n, S_1)+ & w_2\bar{\gamma}(S_n, S_2)+ & w_3\bar{\gamma}(S_n, S_3)+ & \dots + & w_n\bar{\gamma}(S_n, S_n)+ & \lambda = \bar{\gamma}(S_n, A) \\
 w_1 & w_2 & w_3 & \dots & w_n & = 1
 \end{array}
 \tag{2.2.7}$$

Considere la primera ecuación, el término independiente requiere meramente los valores del promedio del semi-variograma entre la muestra 1 y el área desconocida. Los términos del lado izquierdo contiene los $n + 1$ elementos desconocidos, los pesos w_i y los valores del promedio del semi-variograma entre la muestra 1 y cada una de las otras muestras a su vez.

La segunda ecuación es idéntica a la primera sólo que es la muestra 2 la que está presente a lo largo de la ecuación. El tercero tiene la muestra 3 de la misma manera a lo largo de la ecuación, y así sucesivamente hasta la ecuación n con S_n . Finalmente, se tiene la condición necesaria para la suma de los pesos. La solución a este juego de ecuaciones producirá un juego de pesos w_i que dan el Mejor Estimador Insesgado Lineal (BLUE).

La varianza del estimador de Kriging podría ser encontrada por la substitución de los pesos en la ecuación general de la varianza de la estimación (2.2.5). Para la estimación del valor “desconocido” T^* del área o punto de interés A , se sustituyen los pesos w_i en la ecuación 2.2.3.

2.3. Conclusiones parciales

La interpolación basada en Splines Cúbicos es un método de interpolación por tramos que utiliza un polinomio de grado 3. La utilización de un polinomio de este tipo en lugar de un polinomio general de mayor grado n (en el caso de interpolación global) da una mejoría en la

complejidad de los cálculos y permite imponer condiciones adicionales a la función interpoladora: la existencia y continuidad en cada punto de derivada primera, segunda, etcétera, hasta donde sea posible, lo cual implica lograr el mayor grado posible de “suavidad” en la función de interpolación. La formulación matemática presentada en el capítulo demostró como se resolvió el problema de hallar una función “suave” que permitiera solucionar el problema práctico de aproximar una curva plana, suave, que pase por un conjunto dado de $n + 1$ puntos.

La interpolación basada en Kriging usa el variograma como medio para determinar la magnitud de influencia de observaciones vecinas al predecir los valores en las locaciones (puntos) desconocidas. En Kriging, además de tener en cuenta el error, se toma en cuenta la variabilidad espacial global del fenómeno bajo estudio y no solo la variación sistemática o determinista, esto es lo que le permite a Kriging obtener el mejor estimador lineal insesgado (BLUE).

Capítulo 3

Métodos de diseño y modelación del Software

3.1. Introducción

Cada proyecto software es diferente, esta singularidad proviene de las diferencias en el tipo de sistema, en el cliente, en la organización de desarrollo y en la tecnología (Jacobson, Booch, & Rumbaugh, 2000). Las herramientas y principios empleadas para el diseño y la modelación de la solución al problema de investigación, y justificación para su uso, se presentan a continuación. Más adelante, se presenta una descripción de la solución propuesta.

3.2. Herramientas y Principios

3.2.1. La tecnología y lenguaje de programación en MATLAB

MATLAB (Matrix Laboratory) es un lenguaje de alto rendimiento para la informática técnica. Es un paquete de software que permite hacer matemática y cómputo, visualización, análisis de datos, desarrollo de algoritmos, simulación y modelado, y programar en un ambiente abierto flexible. Ofrece un lenguaje intuitivo a ingenieros, científicos y matemáticos para expresar problemas y sus soluciones matemática y gráficamente (Davis & Sigmon, 2005).

Para el desarrollo de la implementación del sistema, se empleó el ambiente de MATLAB R2007a, sobre plataforma Windows.

3.2.2. Conceptos de programación orientada a objetos (POO)

La programación orientada a objetos (POO) es un método de implementación en que los programas son organizados como colecciones cooperativas de objetos, cada uno de los cuales representan un caso de alguna clase, y en el cual pertenecen a una jerarquía de clases unidos vía las relaciones de herencia (Booch, 1994).

La POO provee los conocimientos necesarios para el diseño e implementación de sistemas o clases del sistema, facilitando la codificación, corrección, extensibilidad y reusabilidad de las aplicaciones (Stroustrup, 1988).

3.2.3. La programación en COM y controles de Active X

Modelo de Objeto de Componente (COM) es un juego de herramientas y tecnologías orientadas a objetos que les permiten a diseñadores del software que integren los componentes de tipo aplicación-específicos de los diferentes vendedores en su propia aplicación a través del uso de Interfases. Un interfaz proporciona el acceso a las propiedades y métodos de un objeto de COM. Por ejemplo, usando COM, un diseñador puede escoger un componente de acceso de base de datos por un vendedor, un componente del gráfico comercial por otro, y puede integrar éstos en un paquete del análisis matemático producido por un tercero.

COM provee un Framework para integrar los componentes de software reusables en una aplicación sin la necesidad de recompilar la aplicación entera. Dado que se llevan a cabo los componentes con el código compilado, el código fuente puede escribirse en cualquiera de los muchos lenguajes de programación que apoyan COM.

Un control de ActiveX es un componente que tiene un interfaz del usuario, permitiéndole que responda a las acciones tomadas por el usuario. Por ejemplo, una hoja de cálculo de Microsoft Excel puede ejecutarse como un control. Un control corre en el espacio de proceso de su aplicación cliente. Se dice que el cliente es un recipiente del control porque contiene el control. La aplicación del cliente puede acceder directamente a los métodos y propiedades de un control y el control puede comunicar los eventos de input del usuario al cliente.

3.2.4. Ejecutables de MATLAB (MEX-Executables) creado con Fortran, C/C++

Son códigos hechos en otros lenguajes de programación y compilados como ejecutables o DLL (Dynamic Link Libraries - bibliotecas de enlaces dinámicos) para funcionar en el ambiente de MATLAB. En la figura se resume el proceso de hacer un MEX de C/C++ [ver Anexo B.1].

3.2.5. Modelación en UML (Unified Modeling Language)

El Lenguaje Unificado de construcción de Modelos es un lenguaje gráfico para visualizar, especificar, construir y documentar los artefactos de un sistema de software (Booch, Rumbaugh, & Jacobson, 2000). Los modelos son realizados para comunicar la estructura deseada y el comportamiento de un sistema; para visualizar y controlar la arquitectura del sistema; para comprender mejor el sistema que se esté construyendo, muchas veces hallando oportunidades para la simplificación, la reutilización y para controlar el riesgo (ver la Figura 3.1) (Alarcón, 2000).

A través del modelado logramos 4 objetivos:

- ▷ Visualizar como es o como queremos que sea el sistema
- ▷ Especificar la estructura o el comportamiento del sistema
- ▷ Proporcionar plantillas que nos guíen en la construcción del sistema
- ▷ Documentar las decisiones que hemos tomado

Construimos modelos de sistemas complejos porque no podemos comprender el sistema en su totalidad (Booch *et al.*, 2000).

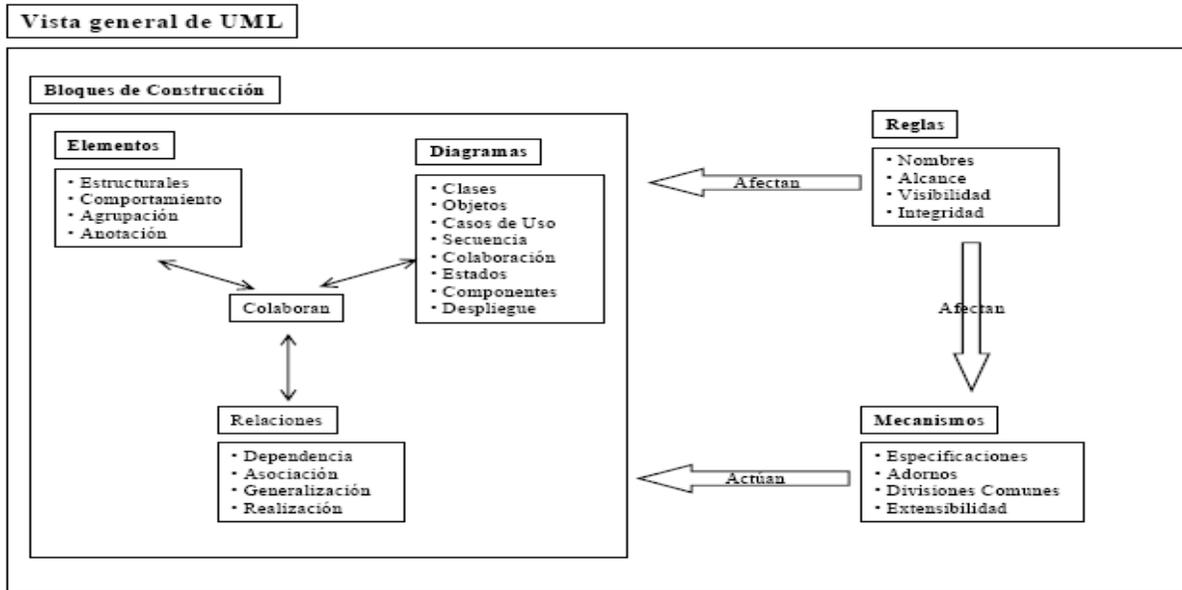


Figura 3.1: Vista general de los elementos de UML.

3.2.6. La metodología RUP (Rational Unified Process)

RUP es el resultado de más de 30 años de experiencia y uso práctico de Proceso Unificado. Su desarrollo como producto sigue en camino desde el Proceso Objectory (primera publicación en 1987) pasando por el Proceso Objectory de Rational (publicado en 1997) hasta el Proceso Unificado de Rational (publicado en 1998) (Jacobson *et al.*, 2000). Según Jacobson *et al.* (2000), el Proceso Unificado es un proceso de desarrollo de software. Un proceso de desarrollo de software es el conjunto de actividades necesarias para transformar los requisitos de un usuario en un sistema software. Sin embargo, el Proceso Unificado es más que un simple proceso; es un marco de trabajo genérico que puede especializarse para una gran variedad de sistemas software, para diferentes áreas de aplicación, diferentes tipos de organizaciones, diferentes niveles de aptitud y diferentes tamaños de proyecto. El objetivo del Proceso Unificado es guiar a los desarrolladores en la implementación y distribución eficiente de sistemas que se ajusten a las necesidades de los clientes.

El Proceso Unificado utiliza el Lenguaje Unificado de Modelado (Unified Modeling Language, UML) para preparar todos los esquemas de un sistema software. De hecho, UML es una parte esencial del Proceso Unificado -sus desarrollos fueron paralelos.

No obstante, los verdaderos aspectos definatorios del Proceso Unificado se resumen en tres frases clave -dirigido por casos de uso, centrado en la arquitectura, e iterativo e incremental. Esto es lo que hace único al Proceso Unificado (Jacobson *et al.*, 2000).

3.2.7. Rational Rose 2003

Para apoyar el trabajo con la metodología RUP ha sido desarrollada por la Compañía norteamericana Rational Corporation en el año 2000 la herramienta de tipo CASE (Computer Assisted Software Engineering) nombrada Rational Rose. Esta herramienta integra todos los elementos que propone la metodología para cubrir el ciclo de vida de un proyecto.

3.2.8. Adobe Photoshop CS3

Aplicación informática que se ha convertido, casi desde sus comienzos, en el estándar mundial de edición y retoque de imágenes; pero también se usa extensivamente en multitud de disciplinas del campo del diseño y fotografía, como diseño Web, composición de imágenes bitmap, estilismo digital, fotocomposición, edición y grafismos de vídeo y básicamente en cualquier actividad que requiera el tratamiento de imágenes digitales.

3.2.9. Microsoft Excel

Una herramienta de Microsoft que brinda las posibilidades de entrar, presentar, explorar y analizar datos (Etheridge, 2007).

3.2.10. Borland C++ Builder 2007

El Borland C++ Builder, es un ambiente de desarrollo rápido de aplicaciones (RAD) muy flexible. Dispone de un entorno visual de fácil uso. Permite la importación de código C++ existente. Posee una gran cantidad de clases y objetos reusables. Es una herramienta de propósito general con una gran documentación disponible (Schildt & Guntle, 2001).

3.3. Modelado de diferentes vistas de una Arquitectura

Según Kruchten (2000), una vista arquitectónica es una descripción simplificada (una abstracción) de un sistema de una perspectiva particular o posición ventajosa, cubriendo preocupaciones particulares y omitiendo entidades que no son pertinentes a esta perspectiva.

Para modelar un sistema desde diferentes vistas fue necesario:

- ▷ Decidir que vistas se necesitan para expresar mejor la arquitectura del sistema e identificar los riesgos técnicos del proyecto.
- ▷ Decidir para cada una de estas vistas qué artefactos se necesitan crear para capturar los detalles esenciales y la relación entre ellos.
- ▷ Decidir el Comportamiento, cómo se especifica las colaboraciones entre esos componentes.
- ▷ Decidir el estilo arquitectónico que guía esta organización: elementos estáticos y dinámicos y sus interfases, sus colaboraciones y su composición.

En el contexto del software hubo tres vistas complementarias que fueron las más importantes para visualizar, especificar, construir y documentar la arquitectura.

- ▷ Vista de casos de uso.
- ▷ Vista de diseño.
- ▷ Vista de implementación.

Cada una de estas vistas involucra modelado estructural (cosas estáticas) y modelado de comportamiento (cosas dinámicas) (ver Figura 3.2). Juntas, estas diferentes vistas capturaran las decisiones más importantes sobre el sistema. Individualmente, cada una de estas vistas permitió centrar la atención en una perspectiva del sistema para poder razonar con claridad sobre las decisiones (Jacobson *et al.*, 2000).

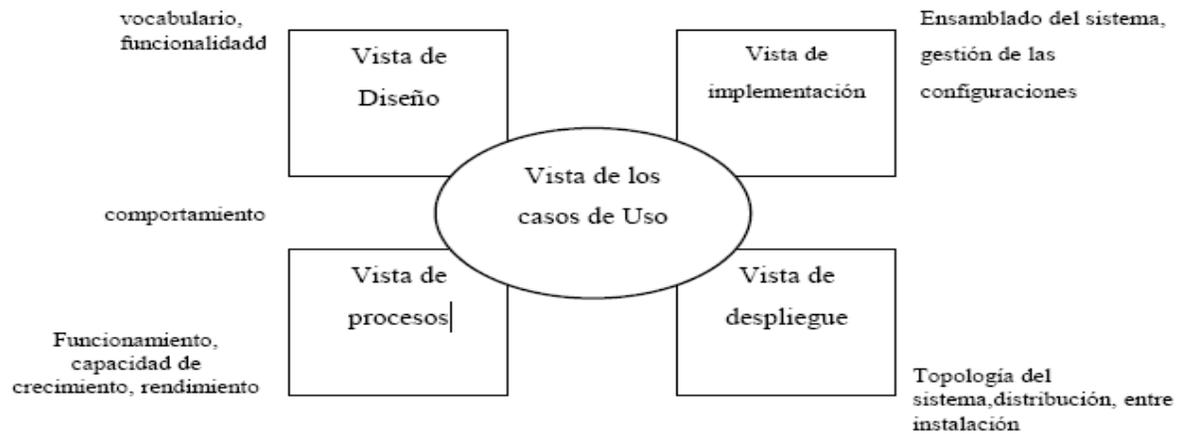


Figura 3.2: Modelado de la arquitectura de un sistema.

Los procedimientos empleados en el diseño de la arquitectura basada en RUP para el sistema de Interpolación de Datos Espaciales Ambientales (IDEA) son expuestos en breve. Entre los múltiples diagramas de UML se emplearon:

- ▷ Diagramas de clases
- ▷ Diagrama de casos de usos
- ▷ Diagrama de componentes

Todos estos diagramas fueron empleados en la concepción de un prototipo que justifique la validez de la idea a defender.

Diseño del Diagrama de clases

Se identificaron varias abstracciones para describir el problema y la solución. Para cada abstracción se identificaron sus correspondientes responsabilidades y se trató de realizar un buen reparto de responsabilidades entre todas las abstracciones. Se proporcionaron los atributos y operaciones necesarias para cumplir las responsabilidades de cada abstracción (Boggs & Boggs, 2002).

Diseño del Diagrama de casos de uso

Para la modelación del contexto del sistema se identificaron los actores en torno al sistema, teniendo en cuenta el comportamiento que cada actor espera del sistema o requiere que este le proporcione.

Se le proporcionó un estereotipo a cada actor para una mejor comprensión de rol con respecto al sistema (Boggs & Boggs, 2002).

3.4. Modelo de Dominio

Según Jacobson *et al.* (2000), hay diferentes puntos de partida para la captura de requisitos. En algunas ocasiones, se comienza haciendo un modelo del negocio, o se comienza con un modelo del negocio que ya está en desarrollo por parte de alguna otra empresa. En otros casos, el software es un sistema que no da soporte directamente al negocio. Entonces se puede tener como entrada un modelo de objetos sencillo, como un modelo del dominio.

El modelo del dominio se utilizó para capturar los tipos más importantes de objetos (clases) en el contexto del sistema. Los objetos del dominio representan las “cosas” que existen o los eventos que suceden en el entorno en el que trabaja el sistema (Jacobson, Christerson, & Övergaard, 1993).

Las clases del dominio aparecieron en dos formas:

- ▷ Objetos del negocio que representan cosas que se manipulan en el negocio.
- ▷ Objetos del mundo real y conceptos de los que el sistema debe hacer un seguimiento.

El modelo del dominio se describió mediante diagramas de UML (especialmente mediante diagramas de clases). Estos diagramas muestran a los clientes, usuarios, revisores y a otros desarrolladores las clases del dominio y cómo se relacionan unas con otras mediante asociaciones

3.4.1. Representación del Modelo de Dominio

Se obtuvieron varias vistas de importancia reflejadas en el diagrama de clases del modelo de dominio perteneciente al módulo de IDEA mostrado en la figura 1, y se definen las entidades y clases relacionadas a continuación:

- ▷ Datos

- ▷ Modelo interpolado
- ▷ Modelo Kriging
- ▷ Modelo Validación
- ▷ Modelo variograma
- ▷ Científico

3.4.2. Diagrama de Clases de Dominio

En la figura 3.3 son mostradas las relaciones existentes entre las clases más importantes en el entorno en el que trabaja el sistema. El sistema utiliza juegos de datos y ocasionalmente un modelo variograma para crear un modelo de interpolación y a través de un modelo de validación se chequear la validez del modelo interpolado.

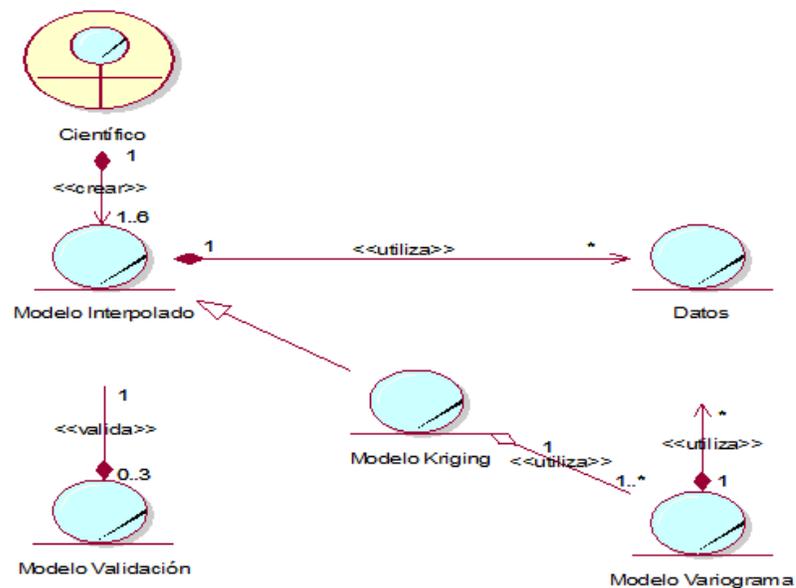


Figura 3.3: Conjunto de clases relacionadas en el Modelo de Dominio.

3.4.3. Descripción del Modelo de Dominio

Un científico utiliza un juego de datos de experimentos y muestras tomadas en el campo de estudio para crear modelos a través de un asistente computarizado. Un juego de datos posee atributos como las variables espaciales X , Y , Z que representen las longitudes, latitudes y valores medidas respectivamente. Un modelo interpolado es el resultado de aplicar una serie de pasos de algoritmos de interpolación sobre los datos. Un modelo de validación es el resultado de aplicar técnicas estadísticas sobre los modelos interpolados. El modelo de validación posee atributos como *RMSE* (Root Mean Square Error - La raíz del error cuadrático medio), *Mean* (media), *Variance* (varianza), *Standard deviation* (desviación estándar). El resultado de aplicar un modelo de variograma sobre los datos sirve como una entrada para el modelo interpolado. Un modelo de variograma posee atributos como la media de la distancia entre puntos de datos (*MeanLag*) y una matriz del variograma experimental (*ExperiVario*).

3.4.4. Reglas del dominio a considerar

Las reglas de negocio regulan y describen las principales políticas o condiciones que deben cumplirse para el adecuado funcionamiento del negocio (Kruchten, 2000). Asimismo, se puede decir que las reglas de dominio realizan un papel similar. A continuación se presentan las que fueron identificadas:

- ▷ Hay que tener datos presentes para crear modelos.
- ▷ Hay que tener un asistente computarizado para crear modelos.
- ▷ Todos los datos utilizados para crear modelos tienen que ser datos espaciales, o sea, que este en la forma de tres variables X , Y , Z que representen los longitudes, latitudes y valores medidas respectivamente.
- ▷ Hay que realizar un modelo interpolado antes de realizar un modelo validación.
- ▷ Hay que tener un modelo variograma antes de realizar un modelo Kriging

3.5. Modelo de Sistema

El esfuerzo principal en la fase de requisitos fue el desarrollo de un modelo del sistema que se iba a construir, y la utilización de los casos de uso fue una forma adecuada de crear este modelo (Jacobson *et al.*, 2000). Se utilizaron los diagramas de casos de uso para modelar la vista de casos de uso del sistema.

3.5.1. Representación del Modelo de Sistema

Requerimientos Funcionales

Los Requerimientos funcionales especificaron las acciones que el sistema debe ser capaz de realizar, sin tomar en consideración ningún tipo de restricción física. Ellos fueron descritos a través del modelo de casos de uso. Por lo tanto los requerimientos funcionales especificaron el comportamiento de entrada y salida del sistema y surgieron de la razón fundamental de la existencia del producto (Jacobson *et al.*, 2000).

Los requerimientos funcionales capturados para el desarrollo del sistema propuesto son los siguientes:

- R1: Entrar datos espaciales desde ficheros de extensión TXT.
- R2: Entrar datos espaciales desde ficheros de extensión XLS (Microsoft Excel).
- R3: Entrar datos espaciales desde ficheros de extensión CSV (Comma Separated Values).
- R4: Entrar datos espaciales desde ficheros de extensión DAT.
- R5: Entrar datos espaciales desde ficheros de extensión MAT (MATLAB).
- R6: Representar gráficamente los datos crudos en la forma de un PLOT.
- R7: Representar gráficamente los datos crudos en la forma de un PLOT3D.
- R8: Representar gráficamente los datos crudos en la forma de un SCATTER.
- R9: Representar gráficamente los datos crudos en la forma de un SCATTER3D.
- R10: Representar gráficamente los datos crudos en la forma de un SURFACE.
- R11: Representar gráficamente los datos crudos en la forma de un SURFACE CONTOUR.
- R12: Representar gráficamente los datos crudos en la forma de un CONTOUR.
- R13: Representar gráficamente los datos crudos en la forma de un CONTOUR3D.

- R14: Representar gráficamente las etiquetas de los valores de los datos crudos.
- R15: Interpolar datos basada en los splines cúbico.
- R16: Interpolar datos basada en los splines bi-armónico.
- R17: Interpolar datos basada en el vecino más cercano (Nearest Neighbor).
- R18: Interpolar datos basada en la interpolación lineal.
- R19: Interpolar datos basada en el peso de la distancia inversa (Inverse Distance Weight).
- R20: Interpolar datos basada en el Kriging.
- R21: Representar gráficamente los datos interpolados en la forma de un CONTOUR.
- R22: Representar gráficamente los datos interpolados en la forma de un CONTOURF.
- R23: Representar gráficamente los datos interpolados en la forma de un PCOLOR.
- R24: Representar gráficamente los datos interpolados en la forma de un MESH.
- R25: Representar gráficamente los datos interpolados en la forma de un MESH.C.
- R26: Representar gráficamente los datos interpolados en la forma de un SURF.
- R27: Representar gráficamente los datos interpolados en la forma de un SURFC.
- R28: Representar gráficamente los datos interpolados en la forma de un SURFCONTOUR3.
- R29: Representar gráficamente los datos interpolados en la forma de un CONTOUR3.
- R30: Representar gráficamente las etiquetas de los valores de los datos interpolados.
- R31: Validar los datos interpolados basada en la Validación Cruzada (Cross-Validation).
- R32: Validar los datos interpolados basada en el Jackknife.
- R33: Validar los datos interpolados basada en el Bootstrap.
- R34: Generar reporte estadística general de la validación de los datos interpolados.
- R35: Generar reporte estadístico de la comparación de dos métodos de interpolación.
- R36: Consultar una ayuda bien detallada sobre las principales opciones del sistema.
- R37: Comparar métodos de interpolación.
- R38: Representar gráficamente la comparación de los interpoladores desde un CONTOUR.
- R39: Representar gráficamente la comparación de los interpoladores desde un CONTOURF.
- R40: Representar gráficamente la comparación de los interpoladores desde un SURF.
- R41: Representar gráficamente la comparación de los interpoladores desde un SURFC.
- R42: Representar gráficamente la comparación de los interpoladores desde un PCOLOR.
- R43: Representar gráficamente la comparación de los interpoladores desde un MESH.
- R44: Representar gráficamente la comparación de los interpoladores desde un MESH.C.
- R45: Representar gráficamente la comparación de los interpoladores desde SURFCONTOUR3.
- R46: Representar gráficamente la comparación de los interpoladores desde un CONTOUR3.
- R47: Crear Modelo de Variograma.
- R48: Representar gráficamente el modelo de variograma.
- R49: Crear Nubes del modelo de variograma (variogram cloud).
- R50: Realizar el proceso de clúster (Clustering).

- R51: Crear región de Máscara de los datos ambientales.
- R52: Exportar datos de interpolación de Splines cúbicos.
- R53: Exportar datos de interpolación de Kriging
- R54: Exportar datos de interpolación de peso de la distancia inversa (Inverse Distance Weight).
- R55: Exportar datos de interpolación del vecino más cercano (Nearest Neighbor).
- R56: Exportar datos de interpolación lineal.
- R57: Exportar datos de interpolación de Splines Bi-armónico.
- R58: Consultar ayuda del manual de usuario.

Requerimientos no funcionales

Los requerimientos no funcionales especificaron las propiedades o cualidades que el producto debe tener. Estas propiedades fueron las características que hacen al producto atractivo, usable, rápido o confiable, por ejemplo, pudiera desearse que el sistema responda dentro de un intervalo de tiempo especificado o que obtenga los resultados de los cálculos con un nivel de precisión dado (Jacobson *et al.*, 2000). En muchos casos los requerimientos no funcionales fueron fundamentales en el éxito del producto.

Los requerimientos no funcionales capturados para el desarrollo del sistema propuesto son los siguientes:

Requerimientos de portabilidad: El software funciona en cualquier ambiente donde este instalado el software MATLAB versión 7 o superior.

Requerimientos de Soporte: El sistema fue realizado con un enfoque flexible, extensible y robusto para propiciar su futuro mejoramiento y la incorporación de otras funcionalidades, módulos u opciones.

Requisitos Legales: La herramienta propuesta responderá a los intereses de la Universidad de Cienfuegos, Centro de Estudio Ambientales de Cienfuegos, Cuba (CEAC) y de la Constitución de la República de Cuba.

Requisitos de Ayuda y Documentación en Línea: Dispondrá de una ayuda bien detallada sobre las principales opciones del sistema y un manual de usuario.

Requerimientos de Confiabilidad: El sistema en casos de fallos debe garantizar que las pérdidas de

información sean mínimas.

Requerimientos de software:

- ▷ Se debe disponer de Windows 98 o superior, para la instalación de la aplicación así como DirectX8.0 o superior para su correcto funcionamiento
- ▷ Se debe tener instalado el ambiente de desarrollo MATLAB versión 7 o superior, sobre plataforma Windows. En caso de no tener instalado el software MATLAB, se debe tener acceso a un servidor que brindará tal servicio.

Requerimientos de hardware: Debido a su empleo en el procesamiento de grandes cantidades de datos, el software depende de un alto nivel de memoria y velocidad de procesador. Para una funcionalidad adecuada se recomienda una computadora de 1 GB de memoria y un procesador Pentium IV con 2.4 GHz y como mínimo una computadora de 512 MB de memoria y un procesador Pentium IV con 2.4 GHz.

Facilidad de Aprendizaje: Una de las razones para construir el sistema fue para facilitar el desarrollo de modelos científicos matemáticos. La construcción de estos modelos lleva consigo muchos algoritmos y fórmulas, y el usuario tiene que tener un amplio conocimiento de los términos técnicos. El sistema facilita el desarrollo de estos modelos ocultando la mayor parte de los algoritmos y formularios, brindando la posibilidad de trabajar en un ambiente automatizado instructivo.

Restricciones en los formatos de fichero: El sistema solamente permite ficheros con extensiones TXT (bloc de notas), DAT, XLS (Microsoft Excel), CSV (Comma Separated Values), MAT (MATLAB).

Restricciones en la plataforma software: El sistema opera con la ayuda del paquete de software MATLAB y es necesario tenerlo instalado para utilizar el sistema.

Restricciones Legales: Debido al injusto bloqueo económico impuesto a Cuba por los Estados Unidos, el acceso a ayuda y documentación desde el sitio oficial del paquete de programa MATLAB (www.MATHWORKS.com) está prohibido desde la isla. Esto causa serías complicaciones en los desarrollos de software, incluso en la extensibilidad del actual software.

3.5.2. Modelo de casos de uso del sistema

Los casos de uso proporcionan un medio intuitivo y sistemático para capturar los *requisitos funcionales* con un énfasis especial en el valor añadido para cada usuario individual o para cada sistema externo (Jacobson *et al.*, 2000). En forma semejante, el modelo de casos de uso permitió la llegada a un acuerdo sobre los requisitos, es decir, sobre las condiciones y posibilidades que debe cumplir el sistema. El modelo de casos de uso sirvió como acuerdo entre clientes y desarrolladores.

Actores del sistema a automatizar

En la siguiente tabla se representa el actor encontrado en el modelado del sistema.

Nombre del actor	Descripción
Usuario	Es el usuario del sistema, que es el encargado de realizar todas funcionalidades asociadas al sistema de interpolación de datos espaciales ambientales. Tiene acceso a todos los requerimientos funcionales del sistema.

Figura 3.4: Actores del Modelo de sistema.

Casos de Usos del sistema

Los casos de uso del sistema quedan representados por:

- ▷ CU1: Entrar Datos Espaciales.
- ▷ CU2: Representar gráficamente los datos crudos.
- ▷ CU3: Representar gráficamente las etiquetas de los valores de los datos crudos.
- ▷ CU4: Interpolar datos.
- ▷ CU5: Representar gráficamente los datos interpolados.
- ▷ CU6: Representar gráficamente las etiquetas de los valores de los datos interpolados.

- ▷ CU7: Validar los datos interpolados.
- ▷ CU8: Generar reporte estadística general.
- ▷ CU9: Generar reporte estadístico comparativo.
- ▷ CU10: Consultar ayuda.
- ▷ CU11: Comparar métodos de interpolación.
- ▷ CU12: Representar gráficamente la comparación de interpoladores.
- ▷ CU13: Crear Modelo de Variograma.
- ▷ CU14: Representar gráficamente el variograma.
- ▷ CU15: Crear Nubes de variograma.
- ▷ CU16: Realizar proceso de clúster.
- ▷ CU17: Crear máscara.
- ▷ CU18: Exportar Datos Interpolados.

Diagrama de casos de Usos

Los diagramas de casos de usos son presentados en forma de Paquetes (Packages). Los paquetes se usan para agrupar los casos de uso basados en una funcionalidad común. En el diagrama de la figura 3.5 se presenta las dependencias entre los paquetes. En la Figura 3.6 se presentan el diagrama de los casos de usos del paquete Interpolación, la Figura 3.7 representa el diagrama del paquete Visualización, la figura 3.8 el paquete Validación y en la figura 3.9 se presenta el paquete CU Generales que representa los casos de usos que no podían ser agrupados en paquetes especializados.

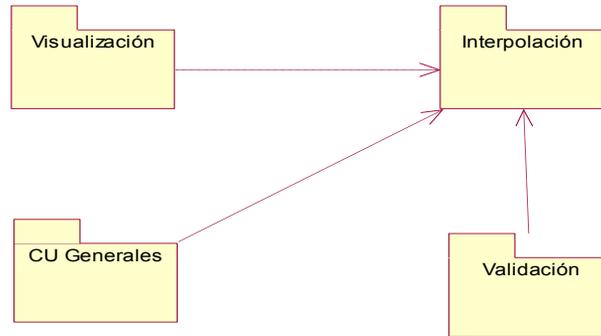


Figura 3.5: Diagrama de casos de usos divididos por Paquetes.

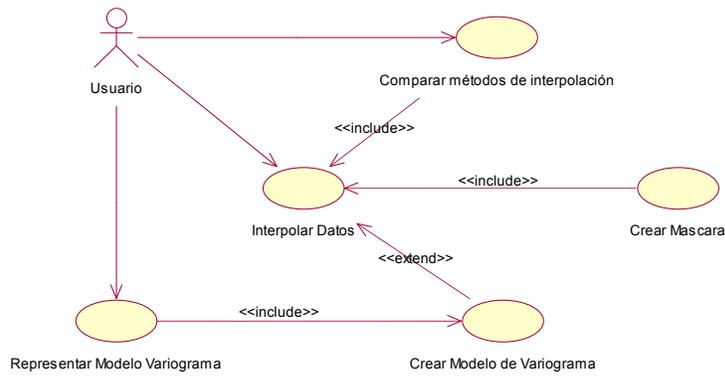


Figura 3.6: Diagrama de casos de usos del Paquete Interpolación.

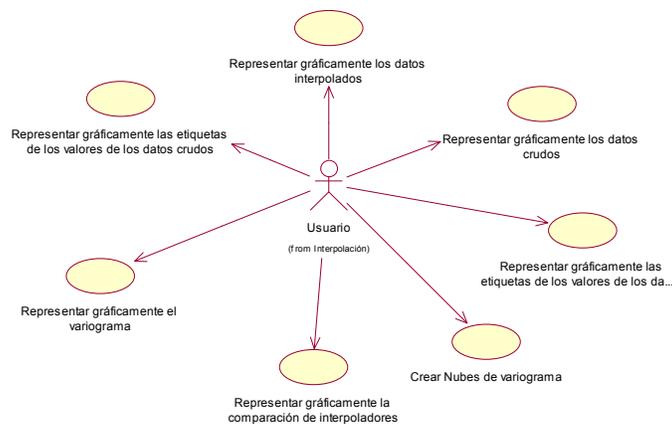


Figura 3.7: Diagrama de casos de usos del Paquete Visualización.

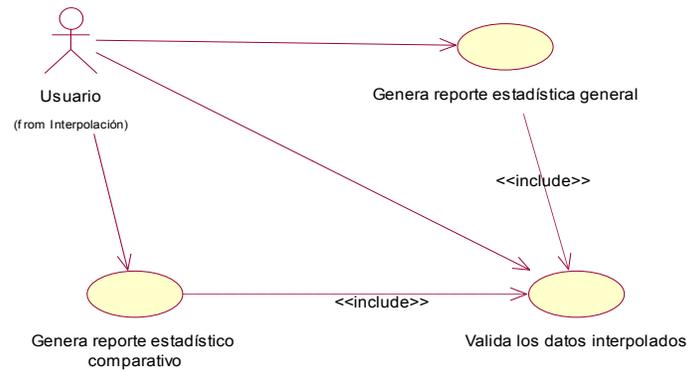


Figura 3.8: Diagrama de casos de usos del Paquete Validación.

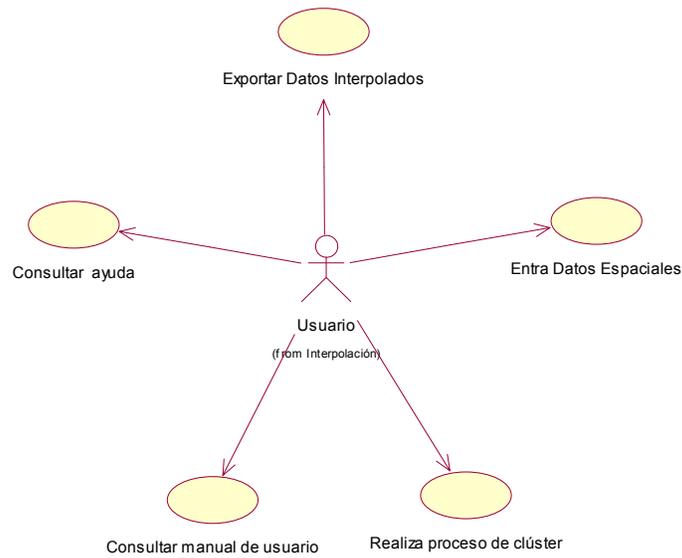


Figura 3.9: Diagrama de los casos de usos del Paquete CU Generales.

Descripción de los casos de usos del sistema

Caso de uso	Entrar Datos Espaciales
Actores:	Usuario
Propósito:	Permite al Usuario entrar los datos espaciales.
Resumen:	El caso de uso se inicia cuando el Usuario decide utilizar el sistema para la interpolación de datos espaciales. El sistema da la posibilidad de entrar estos datos solo si se conforma con la reglas de datos espaciales y el fichero está en una de los formatos permisibles. Para realizar esto se hacen numerosos chequeos y después se preparan los datos eliminando redundancias y se realizan otros requisitos necesarios para la buena utilización de los datos. Si los datos no se conforman, se lanza un mensaje instruyendo al usuario del error y la posibilidad de ver la ayuda. El caso de uso culmina con la aceptación o denegación por parte del sistema de los datos entrados.
Referencias:	R1,R2,R3,R4,R5
Precondiciones:	-
Poscondiciones:	El sistema tiene los datos necesarios para realiza interpolación
Requisitos especiales:	En caso de grandes cantidades de datos, la velocidad del procesador y la tolerancia tiene que ser a niveles elevados para disminuye el tiempo de procesamiento de los datos.
Prototipo:	Ver anexo A.1

Caso de uso	Representar gráficamente los datos crudos
Actores:	Usuario
Propósito:	Permite al Usuario visualizar gráficamente los datos crudos.
Resumen:	El caso de uso se inicia cuando el Usuario decide visualizar el gráfico de los datos crudos. El sistema da la posibilidad de seleccionar entre varias vistas o tipos de gráficas. Después de realizar la selección, el sistema hace un chequeo y deshabilita la opción de mostrar etiquetas de los valores de los puntos de datos crudos si el tipo de gráfico no puede contener etiquetas. El caso de uso se culmina cuando el usuario hace la selección y una presentación visual del gráfico seleccionado es mostrada.
Referencias:	R6,R7,R8,R9,R10,R11,R12,R13
Precondiciones:	Tiene que haber sido entrado datos crudos.
Poscondiciones:	Se crea una gráfica de los datos crudos, se deshabilita la opción de mostrar etiquetas de los valores de los datos crudos.
Requisitos especiales:	-
Prototipo:	Ver anexo A.2

Caso de uso	Representar gráficamente las etiquetas de los valores de los datos crudos.
Actores:	Usuario
Propósito:	Permite al Usuario visualizar los valores reales de los puntos de datos.
Resumen: El caso de uso se inicia cuando el Usuario decide ver los valores de los datos crudos. El usuario hace la selección deseada y hace la actualización de la imagen gráfica de los datos crudos. El sistema responde mostrando el gráfico actualizado con los valores de los datos crudos, culminando el caso de uso.	
Referencias:	R14
Precondiciones:	Debe existir un gráfico de los datos crudos
Poscondiciones:	Un gráfico actualizado con los valores de los datos
Requisitos especiales:	-
Prototipo:	Ver anexo A.3

Caso de uso	Interpolar Datos
Actores:	Usuario
Propósito:	Permite al Usuario interpolar los datos espaciales.
Resumen: El caso de uso se inicia cuando el Usuario decide utilizar el sistema para la interpolación de datos espaciales. El sistema hace un chequeo para ver si el usuario había entrado datos previamente. En casos no afirmativas, se lanza un mensaje instruyendo al usuario de la necesidad de entrar datos y una opción para visualizar la ayuda sobre la preparación de datos y tipos de ficheros permisibles. El usuario tiene la posibilidad de seleccionar entre distintos modelos de interpolación. Después de hacer una selección, el sistema chequear la necesidad de crear un modelo variograma y en casos no afirmativas se lanza un mensaje apropiado. El caso de uso culmina con la selección de la presentación de la ayuda o en el caso que hubo datos previamente entrada, la interpolación de la misma.	
Referencias:	R15,R16,R17,R18,R19,R20
Precondiciones:	Tiene que haber sido entrado dato o en algunas instancias, haber creado un modelo variograma.
Poscondiciones:	La creación de un modelo de interpolación.
Requisitos especiales:	En caso de grandes cantidades de datos, la velocidad del procesador y la tolerancia tiene que ser a niveles elevados para disminuir el tiempo de procesamiento de los datos.
Prototipo:	Ver anexo A.4

Caso de uso	Representar gráficamente los datos interpolados
Actores:	Usuario
Propósito:	Permite al Usuario visualizar gráficamente los datos interpolados.
Resumen: El caso de uso se inicia cuando el Usuario decide visualizar el gráfico de los datos interpolados. El sistema da la posibilidad de seleccionar entre varias vistas o tipos de gráficas. El sistema chequea si hubo datos previamente interpolados. En casos no afirmativa el sistema lanza un mensaje de la necesidad de tener un modelo de interpolación y se da la opción de enlazar hacia la ventana de entrada de datos. El sistema también hace un chequeo y deshabilita la opción de mostrar etiquetas de los valores de los puntos de datos interpolados si el tipo de gráfico no puede soportar etiquetas. El caso de uso se culmina cuando el usuario hace la selección y una presentación visual del gráfico seleccionado es mostrada.	
Referencias:	R21,R22,R23,R24,R25,R26,R27,R28,R29
Precondiciones:	Tiene que haber sido creado un modelo de interpolación
Poscondiciones:	La visualización de distintas vistas de gráficos de un modelo de interpolación.
Requisitos especiales:	-
Prototipo:	Ver anexo A.5

Caso de uso	Representar gráficamente las etiquetas de los valores de los datos interpolados
Actores:	Usuario
Propósito:	Permite al Usuario visualizar los valores interpolados de los puntos de datos.
Resumen: El caso de uso se inicia cuando el Usuario decide ver los valores de los datos interpolados. El usuario hace la selección deseada y hace la actualización de la imagen gráfica de los datos interpolados. El sistema responde mostrando el gráfico actualizado con los valores de los datos interpolados, culminando el caso de uso.	
Referencias:	R30
Precondiciones:	Debe existir un gráfico de los datos interpolados
Poscondiciones:	Un gráfico actualizado con los valores de los datos interpolados
Requisitos especiales:	-
Prototipo:	Ver anexo A.6

Caso de uso	Validar los datos interpolados
Actores:	Usuario
Propósito:	Permite al Usuario analiza la validez de un modelo de interpolación.
Resumen:	El caso de uso se inicia cuando el Usuario decide validar un modelo de interpolación. El sistema da la posibilidad de seleccionar un modelo en particular para chequear la validez del modelo de interpolación. El caso de uso se culmina cuando el usuario hace la selección y una presentación visual del progreso del proceso es mostrada y luego un reporte visual de los indicadores estadísticos de la validación es mostrada.
Referencias:	R31,R32,R33
Precondiciones:	Tiene que haber sido creado un modelo de interpolación
Poscondiciones:	La creación de un modelo de validación
Requisitos especiales:	En caso de grandes cantidades de datos, la velocidad del procesador y la tolerancia tiene que ser a niveles elevados para disminuir el tiempo de procesamiento de los datos.
Prototipo:	Ver anexo A.7

Caso de uso	Generar reporte estadístico general
Actores:	Usuario
Propósito:	Permite al Usuario analiza la validez de todos los modelo de interpolación utilizando un modelo de validación.
Resumen:	El caso de uso se inicia cuando el Usuario decide validar todos los modelos de interpolación. El sistema da la posibilidad de seleccionar un modelo particular de validación para chequear la validez de los modelos de interpolación. El caso de uso se culmina cuando el usuario hace la selección y una presentación visual del progreso del proceso es mostrada y luego un reporte visual de los indicadores estadísticos de la validación es mostrada.
Referencias:	R34
Precondiciones:	Tiene que haber sido entrado datos.
Poscondiciones:	La visualización de indicadores estadísticos de todos los modelos de interpolación
Requisitos especiales:	En caso de grandes cantidades de datos, la velocidad del procesador y la tolerancia tienen que ser a niveles elevados para disminuir el tiempo de procesamiento de los datos.
Prototipo:	Ver anexo A.8

Caso de uso	Generar reporte estadístico comparativo
Actores:	Usuario
Propósito:	Permite al Usuario comparar la validez de dos modelos de interpolación.
Resumen:	El caso de uso se inicia cuando el Usuario decide validar dos modelos de interpolación en forma comparativa. El sistema da la posibilidad de seleccionar un modelo particular de validación para chequear la validez de los modelos de interpolación. El caso de uso se culmina cuando el usuario hace la selección y una presentación visual del progreso del proceso es mostrada y luego un reporte visual de los indicadores estadísticos de la validación es mostrada.
Referencias:	R35
Precondiciones:	Tiene que haber sido creado dos modelos de interpolación.
Poscondiciones:	La visualización de indicadores estadísticas de dos modelos de interpolación
Requisitos especiales:	En caso de grandes cantidades de datos, la velocidad del procesador y la tolerancia tiene que ser a niveles elevados para disminuir el tiempo de procesamiento de los datos.
Prototipo:	Ver anexo A.9

Caso de uso	Consultar ayuda
Actores:	Usuario
Propósito:	Permite al Usuario consultar ayuda sobre las funcionalidades del sistema.
Resumen:	El caso de uso se inicia cuando el Usuario decide consultar la ayuda. El sistema da la posibilidad de seleccionar un tema específico, ver la ayuda desde un página web en forma general, y ver el manual de usuario. El caso de uso se culmina cuando el usuario hace la selección y una presentación visual de la información es mostrada.
Referencias:	R36,58
Precondiciones:	-
Poscondiciones:	La visualización de textos informativos de ayuda.
Requisitos especiales:	-
Prototipo:	Ver anexo A.10

Caso de uso	Comparar métodos de interpolación
Actores:	Usuario
Propósito:	Permite al Usuario comparar dos modelos de interpolación.
Resumen:	El caso de uso se inicia cuando el Usuario decide realizar la comparación de dos métodos de interpolación, uno como principal y otro comparativo. El sistema hace posible que el modelo de interpolación principal en el inicio del caso de uso es siempre el método de interpolación seleccionado en el caso de uso interpola datos. El sistema da la posibilidad de intercambiar la selección del modelo de interpolación principal y modelo comparativo. El usuario tiene la posibilidad de seleccionar entre todas las combinaciones de métodos principales y comparativos. El caso de uso se culmina cuando el usuario hace la selección y una presentación gráfica de la comparación es mostrada.
Referencias:	R37
Precondiciones:	Tiene que haber sido creado un modelo de interpolación.
Poscondiciones:	La visualización de la comparación de dos modelos de interpolación
Requisitos especiales:	-
Prototipo:	Ver anexo A.11

Caso de uso	Representar gráficamente la comparación de interpoladores
Actores:	Usuario
Propósito:	Permite al Usuario visualizar la comparación dos modelos de interpolación.
Resumen:	El caso de uso se inicia cuando el Usuario decide cambiar el tipo de gráfico utilizado para visualizar la comparación de dos métodos de interpolación. El usuario tiene la posibilidad de seleccionar entre distintas vistas de gráficos para la comparación. En el caso que el tipo de gráfico seleccionado no lleve etiquetas, el sistema deshabilita la opción de mostrar etiquetas de los valores de los datos interpolados. El caso de uso se culmina cuando el usuario hace la selección y una presentación gráfica de la comparación es mostrada.
Referencias:	R38,R39,R40,R41,R42,R43,R44,R45,R46
Precondiciones:	Tiene que haber sido hecho la comparación de interpoladores.
Poscondiciones:	La visualización de distintas vistas de gráficos de la comparación de dos modelos de interpolación.
Requisitos especiales:	-
Prototipo:	Ver anexo A.12

Caso de uso	Crear Modelo de Variograma
Actores:	Usuario
Propósito:	Permite al Usuario crear un modelo de variograma.
Resumen: El caso de uso se inicia cuando el Usuario decide crear un modelo de variograma para la interpolación de datos. El sistema hace un chequeo para ver si el usuario había entrado datos previamente. En casos no afirmativos, se lanza un mensaje instruyendo al usuario de la necesidad de entrar datos y una opción para visualizar la ayuda sobre la preparación de datos y tipos de ficheros permisibles. El usuario tiene la posibilidad de seleccionar el tipo de modelo deseado. El caso de uso culmina cuando el usuario hace la selección del modelo.	
Referencias:	R47
Precondiciones:	Tiene que haber sido entrado datos.
Poscondiciones:	La creación de un modelo de variograma
Requisitos especiales:	En caso de grandes cantidades de datos la velocidad del procesador y la tolerancia tiene que ser a niveles elevados para disminuir el tiempo de procesamiento de los datos.
Prototipo:	Ver anexo A.13

Caso de uso	Representar gráficamente el variograma
Actores:	Usuario
Propósito:	Permite al Usuario visualizar un modelo de variograma.
Resumen: El caso de uso se inicia cuando el Usuario decide ver el gráfico de un modelo de variograma para la interpolación de datos. El caso de uso culmina cuando el usuario hace la selección del modelo y una visualización gráfica es mostrada.	
Referencias:	R48
Precondiciones:	Tiene que haber sido creado un modelo de variograma.
Poscondiciones:	La visualización gráfica de un modelo de variograma
Requisitos especiales:	-
Prototipo:	Ver anexo A.14

Caso de uso	Crear Nubes de variograma
Actores:	Usuario
Propósito:	Permite al Usuario visualizar una nube de variograma.
Resumen: El caso de uso se inicia cuando el Usuario decide ver el gráfico de una nube de variograma. El caso de uso culmina cuando el usuario hace la selección y una visualización gráfica es mostrada.	
Referencias:	R49
Precondiciones:	Tiene que haber sido creado un modelo de variograma.
Poscondiciones:	La visualización gráfica de una nube de variograma.
Requisitos especiales:	-
Prototipo:	Ver anexo A.15

Caso de uso	Realiza proceso de clúster
Actores:	Usuario
Propósito:	Permite al Usuario crear y visualizar clústeres de modelos de interpolación.
Resumen: El caso de uso se inicia cuando el Usuario decide crear un clúster del modelo de interpolación. El sistema da la posibilidad de seleccionar entre dos tipos (generales) de análisis de clúster (análisis multi-variado). Para cada tipo de análisis, el usuario tiene la posibilidad de elegir varias técnicas involucradas en este análisis y representarlas a través de sus respectivos gráficos. El caso de uso culmina cuando el usuario hace la selección y una presentación visual del progreso del proceso es mostrada y luego una visualización gráfica del clúster es mostrada.	
Referencias:	R50
Precondiciones:	Tiene que haber sido creado un modelo de interpolación.
Poscondiciones:	La visualización gráfica de un clúster del modelo de interpolación.
Requisitos especiales:	En caso de grandes cantidades de datos, la velocidad del procesador y la tolerancia tiene que ser a niveles elevados para disminuir el tiempo de procesamiento de los datos.
Prototipo:	Ver anexo A.16

Caso de uso	Crear máscara
Actores:	Usuario
Propósito:	Permite al Usuario crear y visualizar una región enmascarada.
Resumen: El caso de uso se inicia cuando el Usuario decide crear una máscara del modelo de interpolación. El sistema da al usuario la posibilidad de entrar los datos que define la región de la máscara El caso de uso culmina cuando el usuario hace la selección y una presentación visual de la gráfica de la región enmascarada es mostrada.	
Referencias:	R51
Precondiciones:	Tiene que haber sido creado un modelo de interpolación.
Poscondiciones:	La visualización gráfica de una región enmascarada del modelo de interpolación.
Requisitos especiales:	En caso de grandes cantidades de datos, la velocidad del procesador y la tolerancia tiene que ser a niveles elevados para disminuir el tiempo de procesamiento de los datos.
Prototipo:	Ver anexo A.17

Caso de uso	Exportar Datos Interpolados
Actores:	Usuario
Propósito:	Permite al Usuario exportar los datos interpolados.
Resumen: El caso de uso se inicia cuando el Usuario decide guardar los datos interpolados del modelo de interpolación. El sistema da al usuario la posibilidad de exportar los datos interpolados en distintos formatos de ficheros. El caso de uso culmina cuando el usuario hace la selección y una presentación visual de la gráfica de la región enmascarada es mostrada.	
Referencias:	R52, R53, R54, R55, R56, R57
Precondiciones:	Tiene que haber sido creado un modelo de interpolación.
Poscondiciones:	Un fichero con los datos del modelo de interpolación es creada.
Requisitos especiales:	-
Prototipo:	Ver anexo A.18

3.6. Construcción del sistema

En este epígrafe se presenta el modelo del diseño para el caso de estudio donde se identifican las clases involucradas y las relaciones que existan entre ellos. Al representar conceptos del mundo real, es importante capturar el estado, las relaciones, y el comportamiento del mundo real. Pero los conceptos de implementación, tales como ocultación de información, eficacia, visibilidad, y métodos, no son conceptos relevantes del mundo real, son conceptos relevantes de diseño (Booch *et al.*, 2000).

3.6.1. Modelo de Diseño

En el modelo de diseño se modeló el sistema y se encontró su forma (incluida la arquitectura) para que soporte todos los requisitos - incluyendo los requisitos no funcionales y otras restricciones que se le suponen (Jacobson *et al.*, 2000). Las funcionalidades del diseño dependen de la clase y los diagramas de interacción. Las clases se definen con la adición de las propiedades, tipos de datos y las operaciones. Cuando esto pasa, el nivel de abstracción de la clase cambia del modo de análisis para el modo de diseñar. Las clases adicionales, principalmente las clases auxiliador y las clases de implementación, fueron agregadas a menudo durante el diseño. El modelo del diseño resultante puede ser trazado directamente en el código. Este fue el enlace entre las abstracciones del negocio (dominio) y las realidades de software.

Diagrama de clases del Diseño

El diagrama de clases de diseño fue uno de los artefactos más significativos para la arquitectura del modelo de diseño (Jacobson *et al.*, 2000). En el diagrama se mostraron las clases de diseño con sus atributos y operaciones, y las relaciones que existen entre clases.

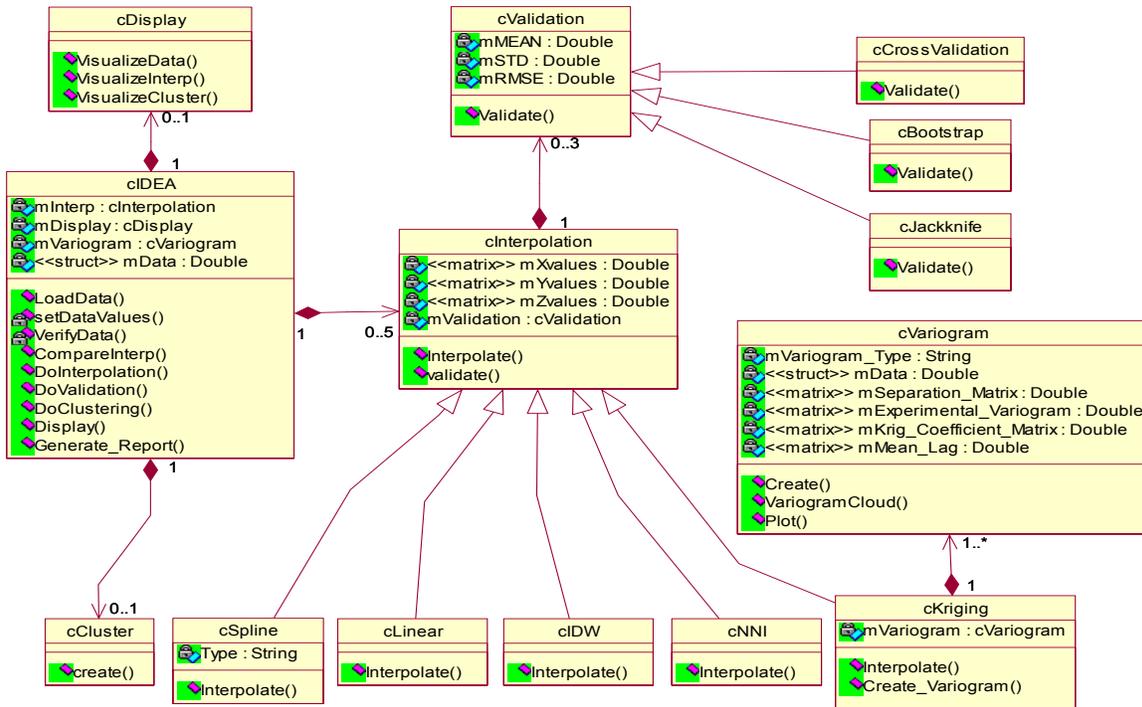


Figura 3.10: Representación del diagrama de clases del Diseño.

3.6.2. Diagrama de implementación

El diagrama de implementación (ver la Figura 3.11) describe como los elementos del modelo de diseño, como las clases, se implementan en términos de componentes. Describe también como se organizan los componentes de acuerdo con los mecanismos de estructuración y modularización disponibles en el entorno de implementación y en el lenguaje o lenguajes de programación utilizados y como dependen los componentes unos de otros (Jacobson *et al.*, 2000).

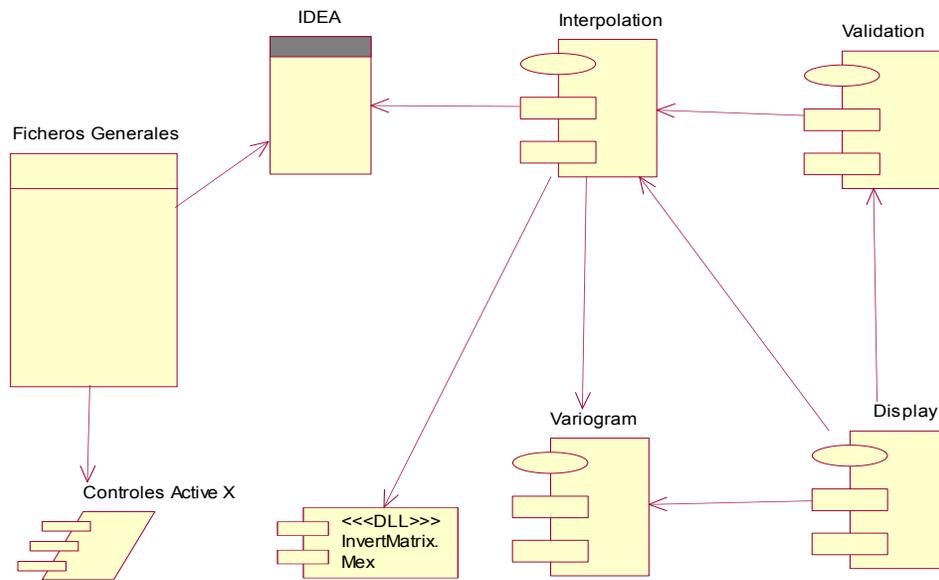


Figura 3.11: Representación del diagrama de componente de implementación.

3.7. Conclusiones parciales

En este capítulo se presentó un modelo de dominio con la arquitectura del modelado estructural (cosas estáticas). El modelado del dominio se contribuyó a una comprensión del problema que se supone que el sistema resuelve en relación a su contexto. A través de un diagrama UML, se representó la primera aproximación de las clases más presentes en el sistema que se propone. En los diagramas de casos de usos se presento el conjunto de secuencias de acciones que el sistema lleva a cabo y que conduce a un resultado observable de interés para los usuarios. Además fueron descritas las reglas que regulan y conducen al buen funcionamiento del sistema. Todo este análisis permitió desarrollar una visión nueva y más clara del problema a resolver.

Capítulo 4

Análisis y Discusión de los Resultados

En este capítulo se presentan los resultados de comparar diferentes métodos de interpolación con la ayuda de dos softwares comerciales (ArcGIS9, SURFER8) y el software desarrollado para esta tesis (IDEA), con el propósito de validar el trabajo hecho en la implementación de los algoritmos de interpolación y su validación. También a través de estas comparaciones, se presentan los problemas existentes en estos paquetes comerciales que generó la necesidad de desarrollar el paquete de software IDEA.

4.1. Caracterización de la muestra

La interpolación de los datos obtenidos del Centro de Estudios Ambientales de Cienfuegos fue utilizada como entrada para crear los gráficos en los paquetes de software. Este juego de datos contiene valores de salinidad tomados de 28 estaciones de la Bahía de Cienfuegos [ver la Figura 4.1]. Estos datos representan el nivel de salinidad presente en el mar en varias locaciones de la Bahía. Este es un juego de datos muy limitado para el cálculo de un variograma experimental. Por consiguiente, se omitieron los variogramas direccionales y sólo los omnidireccionales, (independiente de la dirección), fueron calculados para los modelos de Kriging. En el caso de IDEA y ArcGIS9, se usó un modelo esférico para el variograma y en SURFER8 se utilizó el modelo lineal. En la Tabla 4.1 se presentan el resumen estadísticos de salinidad donde X y Y representan la longitud y latitud respectivamente, de los datos tomados en la Bahía. La variable Z representa la salinidad medida, cuya media es de 35,866 y una desviación estándar de 0,19345, o sea, resulta con un valor de $35,866 \pm 0,19345$ para cada muestra de salinidad tomada en la Bahía un 10 % de

las veces. El valor máximo obtenido es 36,0267 y el mínimo es 34,9644.

	X	Y	Z
Minimum:	-80.5348	22.0723	34.9644
25%-tile:	-80.488	22.0991	35.8571
Median:	-80.4618	22.1252	35.9274
75%-tile:	-80.4435	22.1496	35.9589
Maximum:	-80.4166	22.1709	36.0267
Midrange:	-80.4757	22.1216	35.49555
Range:	0.1182	0.0986	1.0623
Interquartile Range:	0.0445	0.0505	0.1018
Median Abs. Deviation:	0.0238	0.0244	0.0467
Mean:	-80.46760714	22.12304286	35.86623214
Trim Mean (10%):	-80.46698462	22.12315385	35.89474615
Standard Deviation:	0.029015968	0.027972071	0.193448572
Variance:	0.000841926	0.000782437	0.03742235
Coef. of Variation:			0.005393613
Coef. of Skewness:			-3.612263893

Tabla 4.1: Estadísticos de las mediciones de salinidad

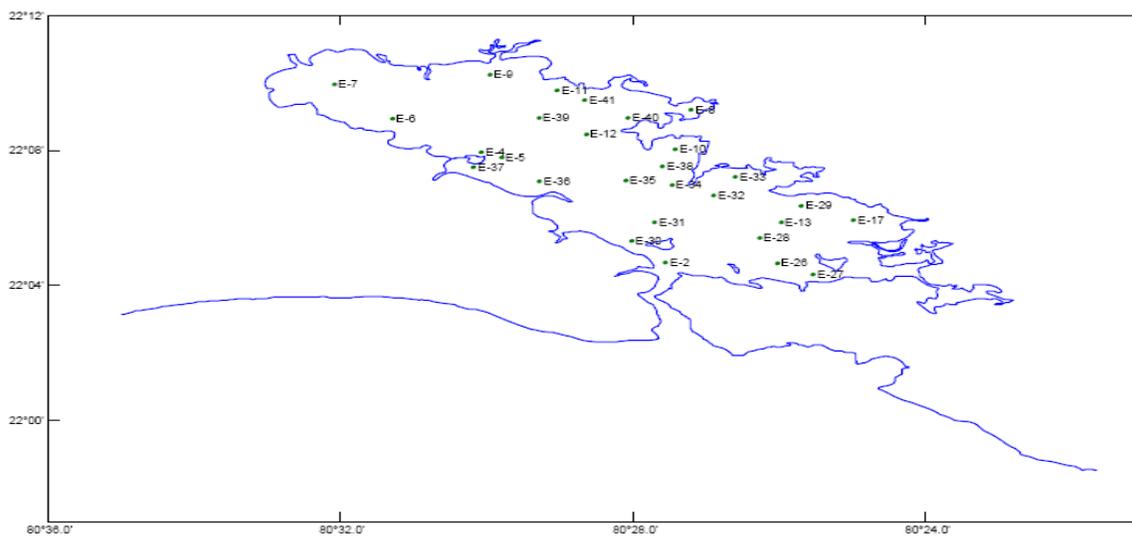


Figura 4.1: Distribución de las estaciones en la Bahía de Cienfuegos

4.2. Caracterización del software

4.2.1. ArcGIS9

En este paquete están disponibles dos métodos de Kriging: Ordinario y Universal. El método Ordinario es el método más usado de los métodos Kriging, mientras el Kriging Universal asume que hay una tendencia principal en los datos (por ejemplo, un viento predominante), y puede ser modelado por una función determinista, por ejemplo, un polinomio. El paquete brinda la posibilidad de seleccionar entre cinco modelos de variogramas [ver la Figura 4.2].

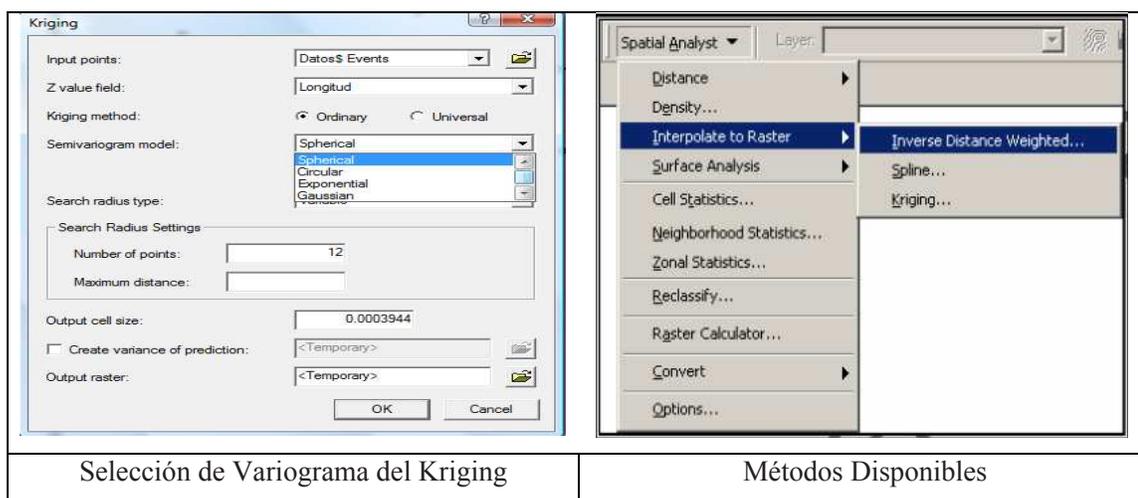


Figura 4.2: Kriging en ArcGIS9

En el caso de interpolación usando Spline, hay dos métodos Spline Regularized y Tension (Regularizado y Tensión) [ver la Figura 4.3]. El método Regularizado crea una superficie suave, de manera gradual cambiando la superficie con los valores que pueden estar fuera de la gama de datos de la muestra. El método de Tensión suaviza la rigidez de la superficie según el carácter del fenómeno modelado. Esto crea la superficie lisa con valores obligados por la gama de datos de la muestra.

También cuenta con la interpolación usando el Método IDW (Gravitacional o Inverso de la Distancia). IDW para el análisis espacial puede ser variable o fijo dependiendo del radio de búsqueda [ver la Figura 4.3]. Para el primero de ellos (fijo) el radio del círculo usado para buscar

entradas de puntos es igual para cada celda interpolada. Para especificar una cantidad mínima se puede asegurar que dentro del radio fijo al menos un número mínimo de puntos de entrada será usado en el cálculo de cada celda interpolada.

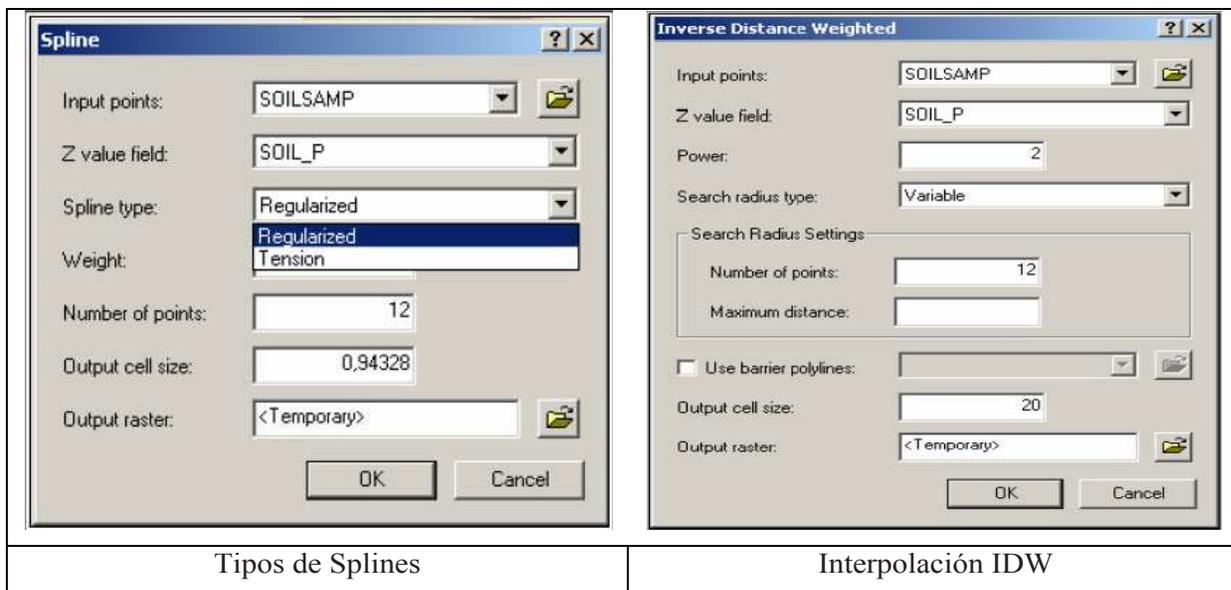


Figura 4.3: Selección del radio de búsqueda en ArcGIS9

4.2.2. SURFER8

Los métodos de gridding (interpolación) en SURFER8 permiten producir contorno exacto, superficie, vector, imagen, y los mapas sombreados de los datos espaciales (XYZ). Los datos pueden ser dispersados al azar en el área del mapa, sin embargo el método de gridding de SURFER8 interpola los datos utilizando una rejilla regularmente espaciada como referencia. En el software existen doce métodos diferentes de gridding [ver la Figura 4.4] brindándole mayor potencia de cálculo al usuario, entre los métodos que tiene implementado aparecen el Método IDW, Kriging y tiene Spline en la forma de una Regresión Polinómica y Polinomio Local como se mencionó en el epígrafe 1.4.4. Cada método de gridding proporciona la funcionalidad para modificar los parámetros, con vista a producir el mapa exacto. Si los datos ya están reunidos en una serie regularmente espaciada, se puede crear un mapa directamente de los datos.

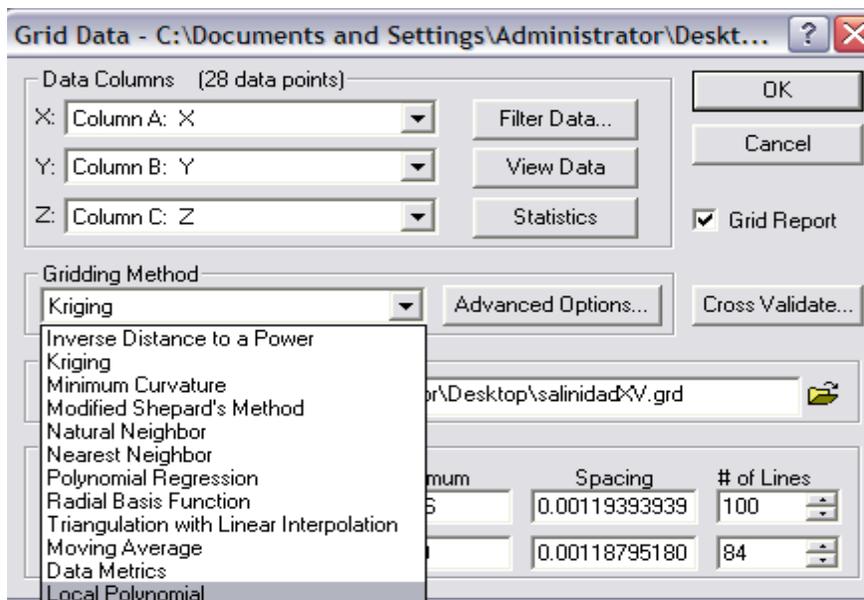


Figura 4.4: Métodos de interpolación en SURFER8

4.2.3. IDEA

IDEA, Versión 1.1 es un programa de interpolación de datos espaciales ambientales sobre la plataforma MATLAB. Con el paquete de software se pueden realizar aproximaciones de datos tomados de investigaciones sobre indicadores en el ambiente, y comparar los métodos de interpolación más usados en esta rama para determinar cuál es más preciso. El programa tiene la ventaja de brindar una interfase gráfica con los algoritmos automatizados, minimizando la necesidad de tener una base técnica fuerte para utilizar dichos algoritmos. Esto permite la operación eficiente por parte del usuario del software para realizar tareas que normalmente conllevan un alto nivel de conocimientos técnicos en la rama de estudios ambientales y la operación de software técnicos como MATLAB.

El paquete de software brinda la posibilidad de interpolar con cinco métodos de interpolación: Splines, Kriging, Inverso de la Distancia Ponderada (IDW), el vecino más cercano (Nearest Neighbor) y la interpolación lineal [ver la Figura 4.5]. El método de Splines incluye los métodos Spline cúbico y Spline bi-armónico, e interpola los datos utilizando una rejilla regularmente espaciada como referencia. La interpolación realizada en Splines y Kriging crea una superficie suave, de manera gradual extrapolando la superficie con los valores que pueden estar fuera de la

gama de datos de la muestra. Esta es una de las razones por las cuales la interpolación realizada en IDEA crear mejores representaciones visuales (mapas) que el ArcGIS9 y SURFER8. Para realizar la interpolación de Kriging el usuario tiene la posibilidad de crear tres modelos de variograma: exponencial, esférico y gaussiano. También, brinda la funcionalidad de validar la exactitud de la interpolación utilizando tres métodos de validación: Cross-validation, Jackknife y Bootstrap.

El paquete tiene una funcionalidad única que permite la creación de una región enmascarada con datos de coordenadas (X, Y) que representa la longitud y latitud de cada punto que forma la región [ver Anexo A.17]. Esta región da la posibilidad de concentrar la interpolación en un área específica minimizando la influencia de los puntos no deseados en el cálculo de la interpolación. Esta es una de las razones por la cual IDEA ha reportado mapas más exactos y datos interpolados más precisos y rápidos.

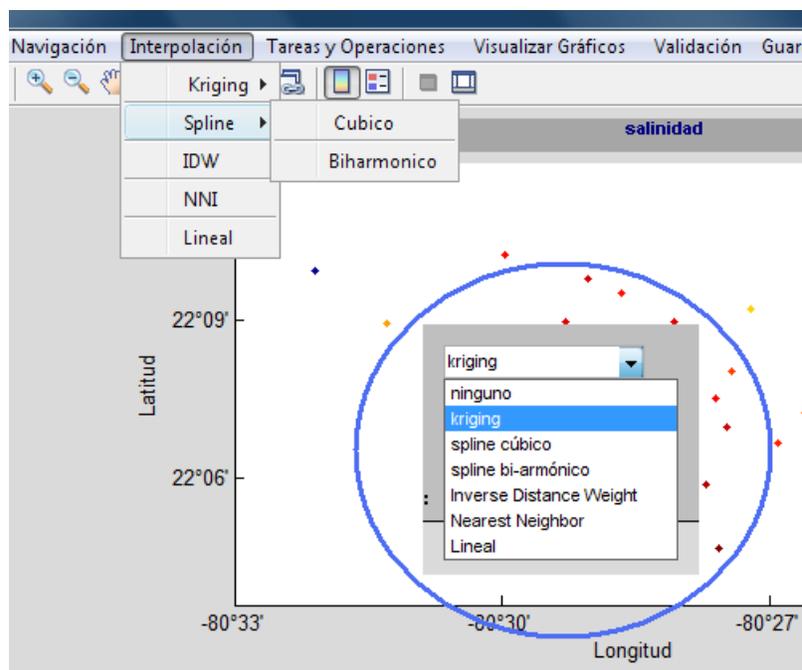


Figura 4.5: Métodos de interpolación en IDEA

4.3. Comparación de los métodos de interpolación en ArcGIS9, SURFER8 e IDEA

Se presentan resultados obtenidos por al menos dos métodos en diferentes mapas para cada paquete de software debido al hecho de que no todos los paquetes de software proporcionan el mismo número y tipo de métodos de interpolación. También debe prestarse atención especial a la interpolación realizada en los paquetes de software como SURFER8 y ArcGIS9 porque ellos realizan una interpolación estricta y no permiten extrapolar los datos, en otros términos, éstos paquetes de software realizan la interpolación de una área producida por los datos, creando una rejilla rectangular grande desde el valor de los puntos extremos norte, sur, este y oeste. Para permitir un rango más amplio (de la región) para la interpolación, se deben agregar puntos de datos auxiliares. Puede ser que los datos agregados no necesariamente reflejen la verdadera naturaleza de los datos a interpolar, pero ayuda a ampliar la región rectangular usada para interpolar el juego de datos. Pero mientras que SURFER8 y ArcGIS9 han sido menos fiables, IDEA ha sido más conveniente para producir mapas que interpolan y extrapolan los puntos de datos dados a la misma vez.

4.3.1. Interpolación con ArcGIS9

En la Figura 4.6 se presentan los mapas generados con la interpolación de Splines y Kriging para ArcGIS9. En los mapas se ve el efecto producido por los métodos de interpolación que crean una región rectangular para los datos extremos. En el mapa de Kriging se ve la transición más suave en el espectro de color desde un área interpolada a otra. Las franjas de colores del mapa indican la tendencia de la salinidad en las aproximaciones alrededor del punto medido y entre puntos vecinos. La amplitud de estas, ofrece información sobre cómo este método utiliza la variabilidad en una vecindad para aproximar un punto desconocido, que no ocurre así en los Splines, ya que este utiliza un polinomio utilizando sólo los puntos medidos en la vecindad, no la varianza.

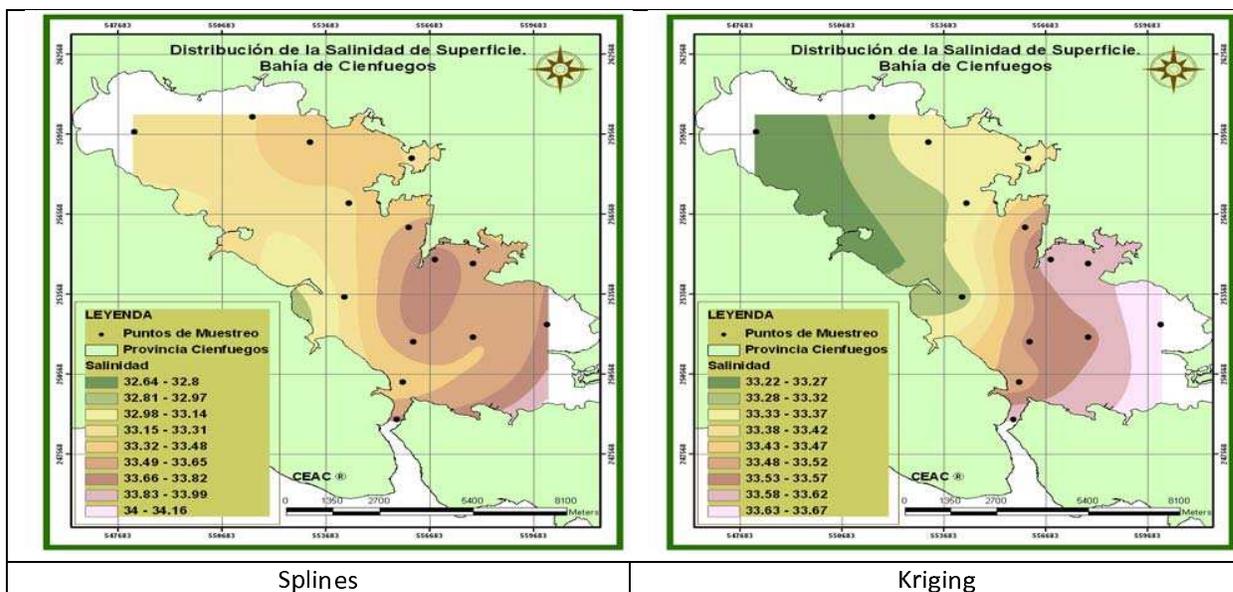


Figura 4.6: Mapas de interpolación de ArcGIS9

4.3.2. Interpolación con SURFER8

En la Figura 4.7 se presentan los resultados de la interpolación haciendo uso del método de Distancia Inversa Ponderada (IDW) y el Vecino Natural (NN) para SURFER8. En el mapa que corresponde al Vecino Natural se ve que el método produce una área interpolada más amplia que el IDW, por lo tanto, brinda más información con respecto a los valores aproximados de los puntos desconocidos en la Bahía.

También, se puede notar en la franja de colores que aprovecha mejor los puntos alrededor brindando mayor información sobre el cambio de la salinidad, así como, la variación al realizar las aproximaciones. Esta diferencia indica que el NN es el método más apropiado para aproximar datos de salinidad de la Bahía.

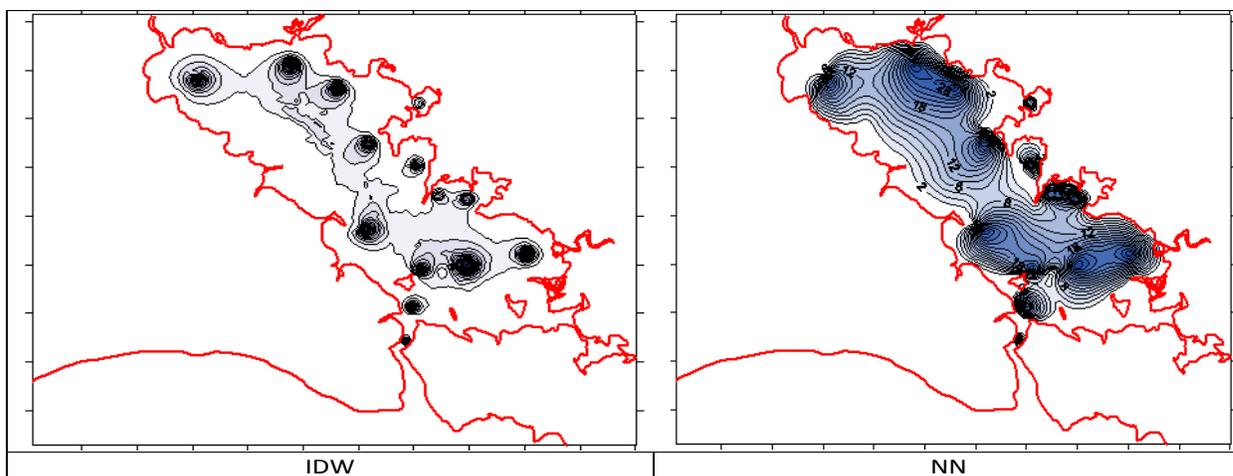


Figura 4.7: Mapas de interpolación de SURFER8

4.3.3. Interpolación con IDEA

Los resultados de Kriging, Splines, IDW y el Vecino más Cercano (Nearest Neighbor) se presentan en la Figura 4.8 para IDEA. En el mapa producido por el método de interpolación Kriging se nota por la franja de colores que la variabilidad utilizada como información para las aproximaciones era la más alta. En el mapa de Kriging se aprecia que el espectro de color desde un punto de interpolación a otro se distingue más fácilmente porque está caracterizada por cambios suaves y sus bordes están bien delimitados que con respecto a los cambios bruscos que se producen en los mapas generados por los métodos Vecino más Cercano e IDW.

Esta transición suave, también vista en el mapa de Splines, indica que estos métodos de interpolación brindan mejor información sobre el valor aproximado del punto desconocido entre un punto medido y otro, además son más apropiados para dar una representación visual de la distribución de salinidad en la Bahía de Cienfuegos.

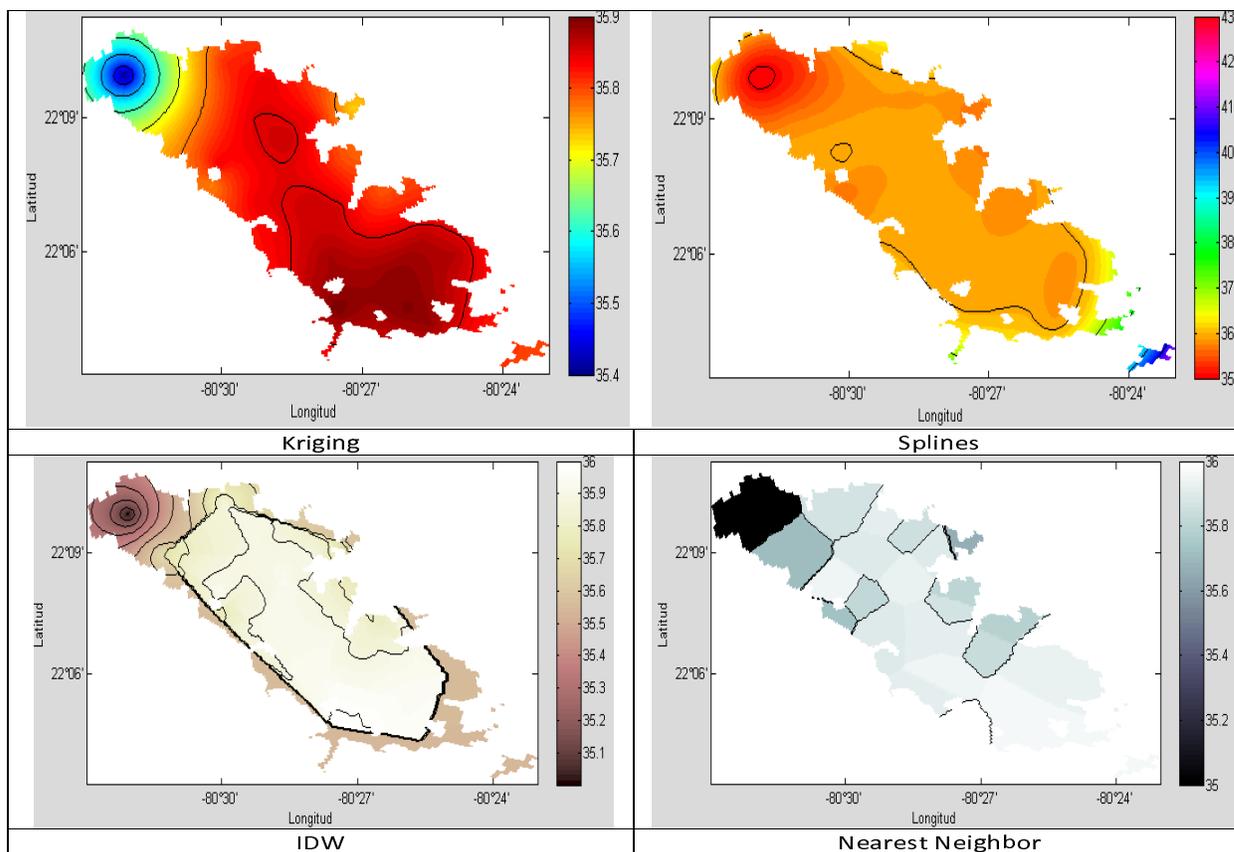


Figura 4.8: Mapas de interpolación de IDEA

En el caso de los mapas generados por la interpolación de IDW y Vecino más Cercano, se puede apreciar el cambio en los contornos de cada área interpolada que es caracterizada por un forma poligonal, ofreciendo poco información sobre la tendencia de la salinidad en la bahía, propiamente en áreas cercanas a los puntos medidos. En IDW, se aprecia que la salinidad que se estima en el área central de la bahía (enmarcada en un polígono) no muestra variación; sin embargo, el mapa del Vecino más Próximo este fenómeno afecta de igual forma pero se puede apreciar que existe una variación tenue pequeñas áreas alrededor del punto medido.

4.4. Validación de los métodos de interpolación

Se realizó una estimación numérica de la precisión de los métodos de interpolación usando la validación cruzada (cross-validation) y Jackknife. Se presentan los parámetros estadísticos

obtenidos de usar estas técnicas, donde MEAN es la media del error de interpolación, VARIANZA y DESVIACION ESTANDAR representan una medida de la distribución del error y RMSE (Root Mean Squared Error) la raíz del error cuadrático medio.

4.4.1. Estadística de la validación en ArcGIS9

En el paquete se realizó la validación cruzada de Kriging, IDW y Splines, los métodos en los cuales es posible realizar este tipo de validación. Según el estudio bibliográfico y análisis del programa, este paquete realiza la validación de modelos de IDW, polinómica (local y global), funciones de bases radiales, Kriging y Cokriging, ofreciendo la raíz del error cuadrático medio (RMSE) y algunos estadísticos descriptivos. Se presenta un resumen del resultado del proceso en la siguiente tabla:

Tipo de Validación	Métodos de Interpolación	MEDIA	VARIANZA	DESVIACION ESTANDAR	RMSE
Cross-validación	Kriging	0.01543	0.0316	0.177833	0.1785
	IDW	0.0293	0.03107	0.17627	0.1787
	Spline Cúbico	0.195284	2.0162	1.419932	1.433

Tabla 4.2: El resultado de validar los métodos de interpolación en ArcGIS9

En la Tabla 4.2 se muestra el RMSE para cada uno de los métodos de interpolación utilizados, en este caso se puede apreciar la diferencia entre los métodos Kriging y IDW respecto al Spline Cúbico, que muestra el peor resultado de RMSE y el error promedio y su varianza en el ajuste de la salinidad es mayor. Kriging, por su parte, proporciona el mejor ajuste en términos del RMSE.

4.4.2. Estadística de la validación en SURFER8

En el paquete de software SURFER8 se realizó una validación de los métodos de interpolación IDW y Nearest Neighbor (vecino más cercano), Splines y Kriging. En la Figura 4.9 (Izquierda) se muestra la interpolación realizada aplicando el Vecino Natural sin aplicar bordes, aquí se puede apreciar como la interpolación no obtiene aproximaciones de toda el área de la Bahía. Aquí se

obtiene un RMSE= 0, 1116 y muestra un coeficiente de Determinación $R^2 = 0,3233$, lo que indica el porciento de la variabilidad de los datos que se logra explicar mediante este modelo

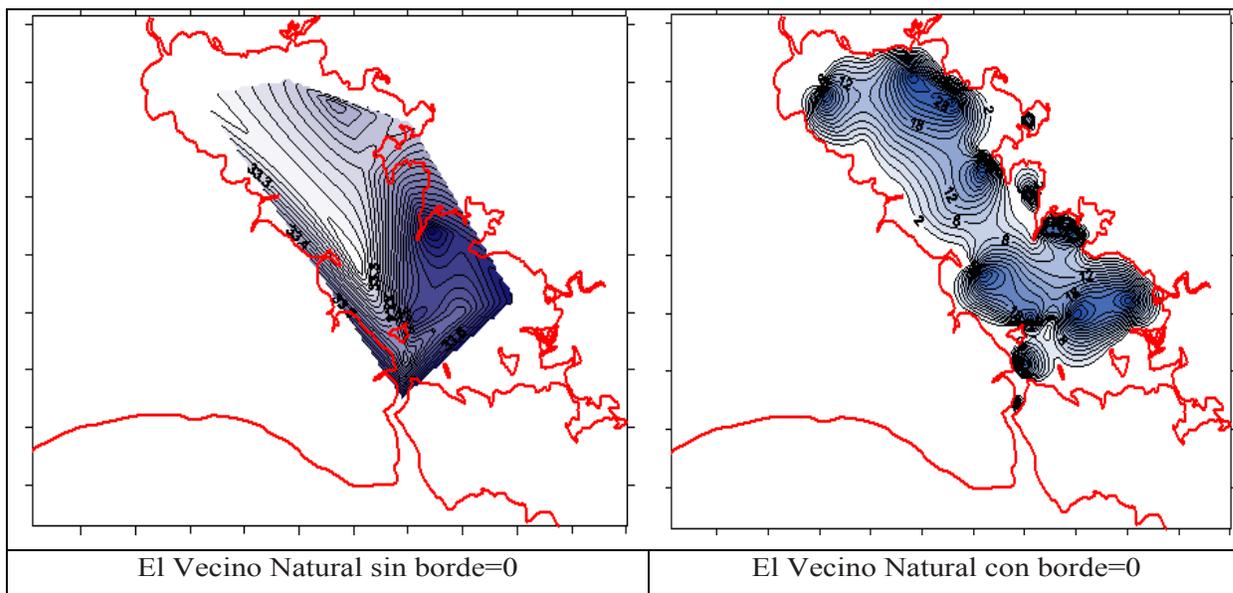


Figura 4.9: Comparación de la interpolación en SURFER8

Para realizar la interpolación en una zona más amplia, fue necesario añadir puntos de datos auxiliares para aumentar la región de interpolación. Para lograrlo, se asignó a la costa de la Bahía el valor 0, pero esto cambió el rango de los valores de la muestra. Como se explicó anteriormente, la adición de nuevos puntos en el modelo a veces produce efectos indeseables. En la Figura 4.9 (Derecha) se muestra como la adición de nuevos puntos mejoró el aspecto visual de la interpolación de datos en SURFER8 pero aumentó el error de aproximación a los datos. La raíz del error cuadrático medio (RMSE) del residuo reportado para el Vecino Natural sin borde fue 0, 1116 y para el Vecino Natural con borde fue (4, 2971).

En la Tabla 4.3 se exponen los resultados de aplicar la técnica de validación cruzada los cuales indican que la interpolación de los datos de salinidad, usando el método del Kriging, es el más exacto, seguido por el método IDW. Otras estadísticas fueron indicadas en el software que brinda información sobre el porcentaje explicado de la variabilidad en los datos mediante el modelo que

asume cada método. En el análisis se calcularon los residuos alrededor de los puntos de validación y fueron utilizados los datos sin borde. En este análisis el Spline resultó el peor, con valor de $RMSE = 0,6835$ y su $R^2 = 0,328$, seguido por IDW con un valor de $0,18173$), seguidas por Kriging ($RMSE = 0,173$) y $R^2 = 0,329$, mientras NN (Vecino más Cercano) reportó un peor valor de $0,2187$ y $R^2 = 0,329$. El valor bajo del coeficiente de determinación múltiple ($R^2 = 0,329$) del residuo reportado para Kriging, combinadas con el valor $RMSE = 0,1730$, verificó que este método es el más preciso en aproximar datos de salinidad de la Bahía. En todos los casos, el R^2 muestra que cualquiera de los métodos utilizados ofrece explicación para el mismo porcentaje de la varianza de los datos (32 % aproximadamente), lo que indica que este parámetro no ofrece información sobre la bondad de ajuste, de los mismos; y a su vez se señala como una debilidad en este programa que realiza la aproximación a través de un modelo lineal. Es lógico esperar un mejor ajuste si se utiliza un polinomio de orden superior.

Tipo de Validación		Métodos de Interpolación	MEDIA	VARIANZA	DESVIACION ESTANDAR	RMSE
Cross-validación	Con borde	Kriging	-0.2177	19.1364	4.37452	4.3799
		IDW	-0.3194	16.8898	4.10972	4.12212
		Nearest Neighbor	-0.31253	21.9089	4.6807	4.6911
		Splines	-0.1773	18.7593	4.3312	4.3348
	Sin borde	Kriging	0.013936	0.029754	0.172495	0.1730569
		IDW	0.02572	0.03236	0.17989	0.18173
		Nearest Neighbor	0.01137	0.04769	0.2184	0.2187
		Splines	-0.09325	0.4584	0.6771	0.6835

Tabla 4.3: El resultado de validar los métodos de interpolación en SURFER8

En la tabla 4.3 es evidente la diferencia en el RMSE de la interpolación realizada con y sin la adición de un borde. Es importante resaltar que la mejora visual obtenida en los métodos de interpolación con borde se obtiene a expensas del deterioro de las medidas descriptivas, entiéndase, la media negativa y una mayor varianza, debido a la inclusión en la muestra de datos que no corresponde

a ningún punto muestral como se explicó anteriormente. Otro parámetro que se ve afectado, es la precisión de aproximación de los métodos de interpolación en SURFER8, lo que hace el software menos fiable.

4.4.3. Estadística de la validación en IDEA

En la Tabla 4.4 se exponen los resultados de aplicar la técnica de validación cruzada los cuales indican que la interpolación de los datos de salinidad usando el método del Kriging, era el más exacto, seguido por el método del Vecino más Cercano. Es importante notar que el Kriging es una interpolación exacta y el objetivo es minimizar el error cuadrático medio. El valor obtenido es un valor aceptable que demuestra que el modelo produjo una buena representación de los datos, y no es sorprendente que el Kriging ha resultado como el más exacto, justificando su reputación de ser el mejor estimador lineal no sesgado (BLUE - por sus siglas en inglés). Sin embargo, es muy sorprendente que el error que resultó por el método del vecino más cercano, sobre todo cuando, comparando por el mapa los gráficos para el vecino más cercano, no mostraron una transformación gradual de las áreas interpoladas cerca de los puntos medios como se ha visto en los métodos de Kriging y splines. Probablemente es un resultado del tamaño del juego de datos usados como entrada, o sea, el número limitado de puntos de datos no puede reflejar la verdadera ventaja de usar los interpoladores exactos.

Las diferencias eran relativamente pequeñas al comparar los métodos de interpolación a través de la técnica de la validación cruzada, sin embargo, los resultados de comparaciones a través de la técnica de Jackknife mostró un alto valor en el error por el método de IDW. El resultado de usar el Jackknife también mostró que los métodos de Kriging y el vecino más cercano eran los más exactos respectivamente.

La semejanza en las tablas (Tabla 4.2, Tabla 4.3 y Tabla 4.4) demuestra que los modelos de interpolación desarrollados en IDEA son significativamente buenos, comparados con los programas profesionales y que la interpolación basada en Kriging es la más apropiada para aproximar los datos de salinidad de la Bahía de Cienfuegos.

Tipo de Validación	Métodos de Interpolación	MEDIA	VARIANZA	DESVIACION ESTANDAR	RMSE
Cross-validación	Nearest Neighbor	-0.01137	0.0494563	0.222388	0.218676
	IDW	0.01817	0.0591151	0.243136	0.239445
	Splines	-0.13726	0.444533	0.666733	0.668952
	Kriging	-0.00717	0.0375961	0.193897	0.190538
Jackknife	Nearest Neighbor	-0.01156	0.0820126	0.286378	0.276203
	IDW	35.8318	0.0735718	0.271241	35.8328
	Splines	0.27785	0.935071	0.966991	0.97236
	Kriging	0.00675	0.0690909	0.262851	0.25338

Tabla 4.4: El resultado de validar los métodos de interpolación en IDEA

4.5. Comparación del error de medición

En la Figura 4.10 se presenta una comparación del error de medición reportado por el proceso de cross-validación. Se utilizó el método de Kriging para realiza el análisis por haber resultado el método que mejor ajusta los datos de salinidad. Se comparó los datos resultados de los modelos predictivos, producidas por la cross-validation de Kriging en cada software.

Se reportó más similitud en los datos producidas por SURFER8 respecto a los datos observados, seguido por IDEA y ArcGIS9 respectivamente. La Figura 4.10 también tiene representada la tendencia lineal de los datos medidos, estimado a través de una recta con coeficiente de determination $R^2 = 0,17$. Este resultado indicó que el modelo no explica bien la variabilidad de los datos y es lógico esperar un mejor ajuste si se utiliza un polinomio de orden superior.

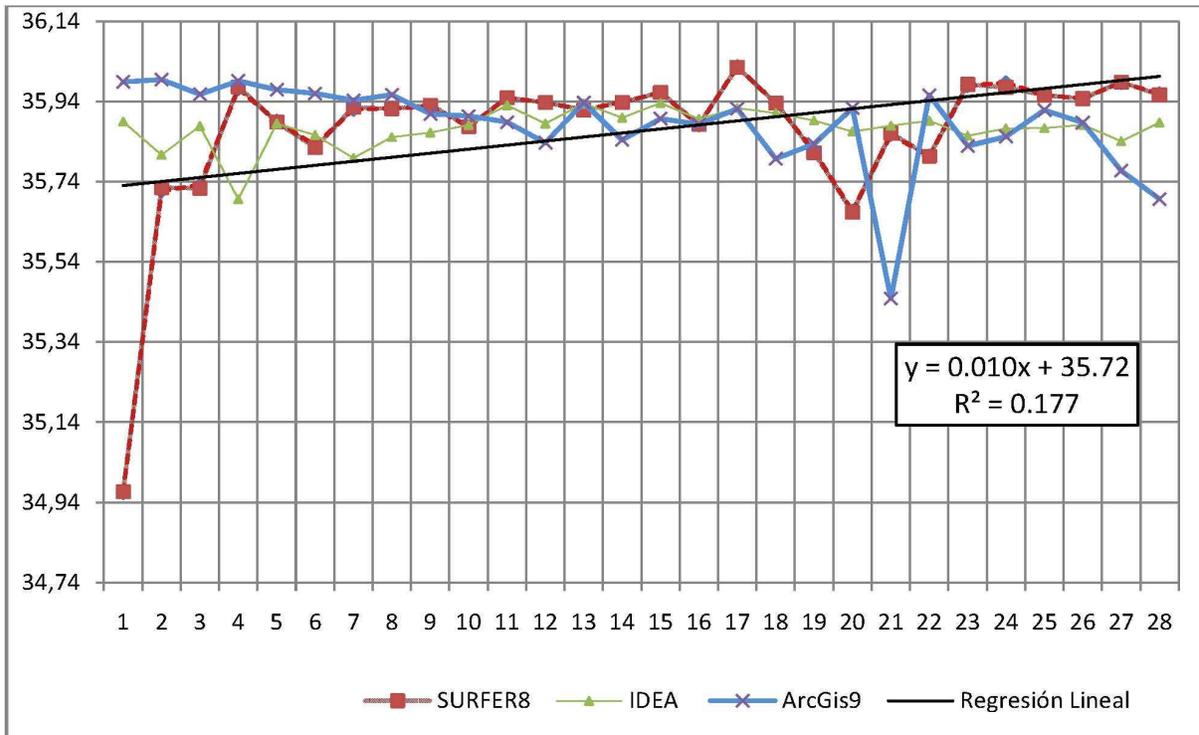


Figura 4.10: Comparación de los modelos predictivos de Crossvalidation

4.6. Costo computacional basado en la arquitectura

En estos experimentos se usó una PC con Pentium Dual Core 1.73 GHz, 1GB de memoria DDR2. El registro de la cantidad de tiempo consumido por el método de interpolación para realizar el cálculo se tomó como una referencia del rendimiento de los algoritmos implementados. Los datos se tomaron de SURFER8 en la forma de tiempo de ejecución (tiempo tomado para realizar la interpolación) en segundos y en el paquete de software de IDEA como tiempo de ejecución y tiempo de CPU. El parámetro de tiempo de CPU indica el tiempo tomado por el procesador (en segundos) para ejecutar el algoritmo de interpolación en el software de IDEA.

En la Tabla 4.5 se muestran los resultados de tiempo consumido por cada aplicación. La interpolación del Vecino más Cercano tiene el menor tiempo de procesamiento (CPU). Este resultado se debe al hecho de que el método es algorítmicamente menos costoso que los otros como Kriging y Splines (en forma de polinomio local de grado 3). El resultado presentado en la Tabla 4.5 da una verdadera representación no sólo de la demanda computacional que se exige del CPU; también refleja su complejidad y comprensibilidad en explorar el juego de datos. En caso

de Kriging (0,98281 segundos), la alta demanda del CPU puede ser atribuida al hecho de que el Kriging explora en forma comprensiva la variabilidad espacial de los datos, primero a través del modelo variograma, y luego realiza las aproximaciones haciendo una minimización del error.

En SURFER8, el tiempo de ejecución era mayor en todos los métodos menos el Kriging comparado

Paquete de Software	Métodos de Interpolación	Tiempo CPU (segundos)	Tiempo de ejecución (segundos)
IDEA	Kriging	0.98281	0.7670
	Splines	0.0468	0.0145
	IDW	0.0165	0.0077
	Nearest Neighbor	0.0156	0.0164
Surfer8	Kriging	-	0.08
	Splines	-	0.51
	IDW	-	0.02
	Nearest Neighbor	-	0.11

Tabla 4.5: Comparación del costo computacional en SURFER8 e IDEA

con IDEA, lo que demuestra que IDEA es más eficaz en realizar las aproximaciones espaciales de los datos de salinidad tomados de la Bahía de Cienfuegos. La mejora en el valor RMSE de Kriging en el Software SURFER8 es el resultado del modelo variograma empleado; este software, utiliza un modelo variograma lineal y según la complejidad de los cálculos, el método de Kriging resulta menos costoso pero produce mapas de peores aspectos visuales y la media del error estimado está más alejada de 0 (dos cifras significativas 0,013936). Para producir mapas de alta calidad visual, en el método de Kriging de IDEA fue utilizado un modelo de variograma esférico que aumentó el costo computacional pero la media del error estimado en este es de $-0,00717$ usando *crossvalidation* y $0,00675$ validando con *Jackknife*.

Es necesario tener en cuenta el hecho de que solo 28 valores fueron tomados para la interpolación de datos sin borde en SURFER8, pero para producir mapas con mayor calidad se requiere añadir más información (datos) lo que significa aumentar el costo computacional. La tabla 4.5 no contiene

datos del tiempo del CPU para SURFER8 debido a la inexistencia de la funcionalidad que permite este tipo de análisis.

4.7. Conclusiones parciales

Se muestran las estimaciones para la salinidad en la Bahía de Cienfuegos en los mapas interpolados por ArcGIS9, SURFER8 e IDEA y la exactitud de las estimaciones se derivó de los resultados de la validación cruzada y el Jackknife. Se usaron cuatro métodos principales de interpolación: La Distancia Inversa Ponderada (IDW), el Vecino más Cercano (NN), los Splines y Kriging. El número de puntos de datos de salinidad fue 28. Es considerado como la cantidad mínima de datos para modelar un juego de datos de entrada desde un modelo de variograma (anisotropic) como en el caso de Kriging.

Se evaluó la calidad de la interpolación de salinidad basado en dos criterios. El primero es de forma gráfica; es la descripción geográfica de la distribución de la salinidad en la Bahía. El segundo es el criterio numérico que representa la estimación de la calidad de la interpolación con respecto al error (RMSE). En este análisis el RMSE más bajo que resultó de la aplicación de la técnica de la validación cruzada fue producido por el Kriging (0,190538) y el más alto fue por el método de Splines (0,668952). Los resultados de aplicar la técnica de Jackknife mostraron que el método de Kriging tenía el error más bajo (0,25338) y más alto fue por el método de la Distancia Inversa Ponderada (IDW) (35,8328).

Según los resultados obtenidos de realizar el Jackknife y la validación cruzada, se puede concluir que la interpolación de Kriging y del Vecino más Cercano son los métodos más apropiados para interpolar los datos de salinidad de la Bahía de Cienfuegos. Sin embargo, una inspección visual de la interpolación desde los mapas muestra que la interpolación de Kriging y de los splines describe mejor el fenómeno geográfico. Es por estas razones se deben usar los resultados de ambos criterios antes de decidir qué método es generalmente más apropiado para interpolar un tipo de datos específico.

En el caso de la salinidad medida en la Bahía de Cienfuegos, la interpolación de Kriging ha resultado como el mejor método de interpolación en el estudio de datos medioambientales espaciales. Pero ningún método es absolutamente mejor que otro, sólo las circunstancias determinan la opción óptima. Primero se debe repasar las características y principios de cada método, las propiedades y distribución espacial de los datos antes de seleccionar un método de interpolación. Por otro lado, el resultado debe evaluarse a partir de las experiencias conscientes de la variable en estudio y su significado físico.

CONCLUSIONES

Como cierre del trabajo se presenta a modo de conclusiones los aspectos más relevantes.

1. Se desarrolla un software (IDEA) basado en la tecnología RUP, sobre MATLAB para la interpolación de variables medioambientales y se encuentra la arquitectura que mejor define la construcción del sistema propuesto.
2. Se realiza un estudio del estado actual sobre la interpolación de datos espaciales medioambientales y se caracterizan los métodos de interpolación más comúnmente usados que son IDW, NN, Kriging y Spline, así como, los métodos de validación, Crossvalidation y Jackknife. Este análisis bibliográfico demuestra que los métodos de Kriging y Splines son los que proporcionan la mejor aproximación para datos ambientales.
3. Se exponen los modelos matemáticos que caracterizan a los métodos de interpolación Kriging y Spline, y se demostró como se resuelve el problema en Spline de hallar una función “suave” que permita solucionar el problema práctica de aproximar una curva plana, suave, que pase por un conjunto dado de $n+1$ puntos. En caso de Kriging, se muestra el uso de la variabilidad espacial global, lo que permite obtener el mejor estimador lineal insesgado (BLUE).
4. Se logra demostrar que el método de Kriging es el que mejor permite interpolar la salinidad de la bahía de Cienfuegos, su implementación en IDEA muestra resultados competitivos respecto a dos programas comerciales utilizados ArcGis9 y SURFER8, este aspecto se determinó basado el tiempo de CPU, la calidad del aspecto visual del mapa interpolado que caracteriza la salinidad y mediante el error (RMSE). El RMSE más bajo resultó para Kriging al validarse con Crossvalidation (0, 190538) y Jackknife (0, 25338).

RECOMENDACIONES

Las recomendaciones de este trabajo son:

1. Implementar métodos iterativos de mayor potencia para el procesamiento de grandes cantidades de datos. Para realizarlo, se propone la programación de los algoritmos de mayor exigencia computacional en un lenguaje más potente como C++ o Java.
2. El desempeño del software tiene una restricción en la dependencia de MATLAB, se propone un diseño de dos capas basado en tecnologías menos dependientes y plataforma independiente como C++ o Java para implementar el interface y el API de MATLAB como DLL (biblioteca de códigos) para implementar el comportamiento.
3. Para acceder a las grandes bases de datos del centro de estudio ambientales (CEAC), se propone la implementación de la tecnología que permite hacer consultas directamente a un sistema de bases de datos, mejorando el rendimiento en el trabajo.
4. La implementación de la funcionalidad que permite al usuario modificar la precisión de realizar las interpolaciones (dar la posibilidad de controlar el tamaño del GRID(Grid size)). El presente trabajo realiza interpolaciones a una precisión fija.
5. La implementación de la funcionalidad que permite al usuario utilizar la información brindada por los métodos de validación como una entrada en los métodos de interpolación, o sea, se deben ajustar los valores aproximados por los métodos de interpolación por el valor de la raíz cuadrada de la media de los errores.

REFERENCIAS BIBLIOGRÁFICAS

- Alarcón, R. (2000). *Diseño orientado a objetos con UML*. Grupo EIDOS Consultaría y Documentación Informática, S.L.
- Balic, D., Velic, J., & Malvic, T. (2008). Selection of the most appropriate interpolation method for sandstone reservoirs in the Kloštar oil and gas field. *Geologia Croatica, 1*, 35.
- Blanco, Álvarez, M., Montenegro, & Gómez, A. (1998). *Matemática Numérica*. Editorial Félix Varela.
- Boggs, W. & Boggs, M. (2002). *Mastering UML with Rational Rose 2002*. Sybex.
- Booch, G. (1994). *Object-Oriented Analysis and Design with Applications* (2nd ed.). Addison Wesley Longman, Inc.
- Booch, G., Rumbaugh, J., & Jacobson, I. (2000). *El Lenguaje Unificado de Modelado*. Addison Wesley.
- Breaz, N. (2004). Numerical experiments with least squares spline estimators in a parametric regression model. *Acta universitatis Apulensis, 8*, 10.
- Castellanos, M. E. G., Sosa, L. P., Moreira, A. R., Maya, H., Pérez, S., León, A. R., & Gómez, M. (2005). Concentración de Arsénico en Macroalgas de la Bahía De Cienfuegos, Cuba. *Rev. Invest. Mar., 26*(1), 21-26.
- Castro-Ruz, F. (2007). *El Diálogo de Civilizaciones*. Oficina de publicaciones del consejo de estado.
- Chegini, Hosseini, E., Mahdian, & Hossein, M. (2001). Survey on application of geostatistical methods for estimation of rainfall in arid and semiarid regions in south west of iran. En *Proceedings of the 6th international conference on geocomputation*. University of Queensland, Brisbane, Australia: GeoComputation CD-ROM. (ISBN:1864995637)
- Childs, C. (2004). Interpolating Surfaces in ArcGIS Spatial Analyst.
- Chiles, J.-P. & Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty*. John Wiley and Sons, Inc.
- Clarke, I. (2001). *Practical Geostatistics*. Geostokos Limited.
- Davis, T. A. & Sigmon, K. (2005). *MATLAB Primer* (7 ed.). Chapman and Hall/CRC.

- Etheridge, D. (2007). *Microsoft Office Excel top 100 simplified tips and tricks*. Wiley Publishing INC.
- Fencík, R. & Vajsáblová, M. (2006). *Parameters of interpolation methods of creation of digital model of landscape*. Unpublished doctoral dissertation, Slovak University of Technology.
- García, S. (2007). Homogeneización e Interpolación de Observaciones de Alta Resolución. *Universidad de Cantabria: España, 73*. (http://www.meteo.unican.es/courses/2007_cursoRegionalizacion_files/1_4_Herrera_Interpolacion.pdf)
- Goovaerts, P. (2006). Geostatistical modeling of the spaces of local, spatial, and response uncertainty for continuous petrophysical properties. En *Stochastic modeling ii*. American Association of Petroleum Geologists.
- Haining, R. (2004). *Spatial data analysis theory and practice*. Cambridge University press.
- Hofstra, N., Haylock, M., Jones, P., & Frei, C. (2008). Comparison of six methods for the interpolation of daily, European climate data. *Journal of Geophysical Research, 113*(D21).
- Jacobson, I., Booch, G., & Rumbaugh, J. (2000). *El Proceso Unificado de Desarrollo De Software*. Addison Wesley.
- Jacobson, I., Christerson, M., & Övergaard, P. J. G. (1993). *Object-Oriented Software Engineering: A Use-Case-Driven Approach*. Addison Wesley.
- Kalivas, P., Triantakou, P., & Kollias, V. (2002). Spatial prediction of two soil properties using topographic information. *Global Nest: the Int. J., 4*(1), 41-49.
- Kiusalaas, J. (2005). *Numerical Methods in Engineering with MATLAB*. Cambridge University Press.
- Krivoruchko, K. & Gribov, A. (2000). A new method for handling the nugget effect in Kriging. Environmental Systems Research Institute, USA.
- Kruchten, P. (2000). *The Rational Unified Process An Introduction* (2nd ed.). Addison Wesley.
- Kurtzman, D. & Kadmon, R. (1999). Mapping of temperature variables in Israel: A comparison of different interpolation methods. *Climate Research, 13*, 33-43.
- Lloyd, C. D. (2007). *Local models for spatial analysis*. CRC Press.
- Mahdian, M. H., Bandarabady, S. R., Sokouti, R., & Banis, Y.Ñ. (2009). Appraisal of the Geostatistical Methods to Estimate Monthly and Annual Temperature. *Journal of Applied Sciences, 9*(1), 128-134.
- Mathews, J. H. & Fink, K. (1999). *Numerical Methods using MATLAB* (3 ed.). Prentice Hall.
- Mathworks. (2009). *Nearest Neighbor Interpolation*. (www.mathworks.com.in/access/helpdesk/help/toolbox/vipblks/ug/f12622.html)
- Miller, H. J. (2004). Tobler's first law and spatial analysis. *Annals of the Association of American Geographers, 94*(2), 284-289.

- NASA. (2009). *Nasa/gsf/met/ersdac/jaros y u.s./japan aster*.
(www.cr.usgs.gov/pub/imswelcome/)
- NC. (1987). *NC 93-02-202/87: Requisitos Higiénicos Sanitarios. Concentración Máximas de Admisible y Altura Mínima de Expulsión y Zonas de Protección Sanitaria* .
- Pecuřová, Z., Parajka, J., & Hruřková, K. (2002). Spatial estimation of snow water equivalent in the mountain basin Bystra. En L. Holko, P. Miklánek, J. Parajka, & Z. K. and (Eds.), *Erb and nefriend proj.5 conf.interdisciplinary approaches in small catchment hydrology: monitoring and research, slovak nc ihp unesco/ih sas* (p. 217-221). Monitoring and Research.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1997). *Numerical Recipes in Fortran 77, The Art of Scientific Computing* (Vol. 1, 2nd ed.). Cambridge University Press.
- Schildt, H. & Guntle, G. (2001). *Borland C++ Builder :The Complete Reference*. Osborne/McGraw-Hill.
- SRTM. (2009). *Shuttle Radar Topography Mission (SRTM)*. (www.jpl.nasa.gov/srtm/)
- Stroustrup, B. (1988). *What is Object-Oriented Programming*. IEEE Software.
- Tolosa, I., Mesa-Albernas, M., & Alonso-Hernandez, C. (2009). Inputs and sources of hydrocarbons in sediments from Cienfuegos Bay, Cuba. *Marine Pollution Bulletin*, 58(11), 1624-1634.
- Tomczak, M. (1998). Spatial Interpolation and its Uncertainty Using Automated Anisotropic Inverse Distance Weighting (IDW) - Cross-Validation/Jackknife Approach. *Journal of Geographic Information and Decision Analysis*, 2(2), 18-30.
- Trauth, M. H. (2006). *MATLAB Recipes for Earth Sciences*. Springer.
- Webster, R. (2007). *Geostatistics for Environmental Scientists* (2nd ed.). John Wiley Sons Ltd.
- Wong, D. W., Yuan, L., & Perlin, S. A. (2004). Comparison of spatial interpolation methods for the estimation of air quality data. *Journal of Exposure Analysis and Environmental Epidemiology*, 14(5), 404-415.

Glosario

Actor:

Cada tipo de usuario se representa mediante un actor. Los actores utilizan el sistema al interactuar con los casos de uso. Todos los actores y casos de uso del sistema forman un modelo de casos de uso.

Arquitectura:

La arquitectura de un sistema es la visión común en la que todos los empleados (desarrolladores y otros usuarios) deben estar de acuerdo, o como poco, deben aceptar. La arquitectura nos da una clara perspectiva del sistema completo, necesaria para controlar el desarrollo.

Centrado en la arquitectura:

El estar centrado en la arquitectura significa que el trabajo de desarrollo se centra en obtener el patrón de la arquitectura que dirigirá la construcción del sistema en las primeras fases, garantizando un progreso continuo no sólo para la versión en curso del producto, sino para la vida entera del mismo.

Clase:

Una descripción de un conjunto de objetos que comparten los mismos atributos, operaciones, relaciones y semántica.

Diagrama:

La presentación gráfica de un conjunto de elementos, usualmente representado como un grafo conectado de vértices (elementos) y arcos (relaciones).

Diagrama de clases:

Incluye clases, interfaces, colaboraciones y las relaciones existentes entre ellas. Estos diagramas describen la vista de diseño estático de un sistema. Describen el vocabulario de un sistema (abstracciones, etc.). Los Diagramas de clase que incluyen clases activas se utilizan para cubrir la vista de procesos estática de un sistema.

Diagramas de casos de uso:

Representa un conjunto de casos de uso y actores y sus relaciones. Estos diagramas son usados para describir la vista de casos de uso estática del sistema. Estos diagramas son especialmente importantes para organizar y modelar el comportamiento de un sistema.

Dirigido por casos de usos:

El estar dirigido por los casos de uso significa que cada fase en el camino al producto final está relacionada con lo que los usuarios hacen realmente. Lleva a los desarrolladores a garantizar que el sistema se ajusta a las necesidades reales del usuario.

Iterativo e incremental:

Con el método de desarrollo iterativo e incremental esta adaptación de los cambios se realiza poco a poco. Dicho de otra forma, dividimos el proyecto en un número de mini proyectos, siendo cada uno de ellos una iteración. Cada iteración tiene todo lo que tiene un proyecto de desarrollo de software: planificación, desarrollo en una serie de flujos de trabajo (requisitos, análisis y diseño, implementación y prueba), y una preparación para la entrega. El ciclo de vida iterativo produce resultados tangibles en forma de versiones internas (aunque preliminares), y cada una de ellas aporta un incremento y demuestra la reducción de los riesgos con los que se relaciona.

Modelo:

Una abstracción de un sistema cerrado semánticamente.

Objeto:

Una manifestación concreta de una abstracción; una entidad sobre la que pueden aplicarse un conjunto de operaciones y que tiene un estado que almacena los efectos de las operaciones; un sinónimo de objeto.

Operación:

La implementación de un servicio que puede ser solicitado por cualquier objeto de la clase para afectar su comportamiento.

Relación:

Una conexión semántica entre elementos.

Sistema:

Una colección de subsistemas organizados para llevar a cabo un propósito específico y descritos por un conjunto de modelos, posiblemente desde distintos puntos de vista.

Vista:

Una proyección de un modelo, la cual es vista desde una perspectiva determinada o punto estratégico y que omite las entidades que no son relevantes para esta perspectiva.

Vista de casos de uso:

Comprende la descripción del comportamiento del sistema tal y como es percibido por los usuarios finales, analistas y encargados de las pruebas y se utilizan los diagramas de casos de uso para capturar los aspectos estáticos mientras que los dinámicos son representados por diagramas de interacción, estados y actividades.

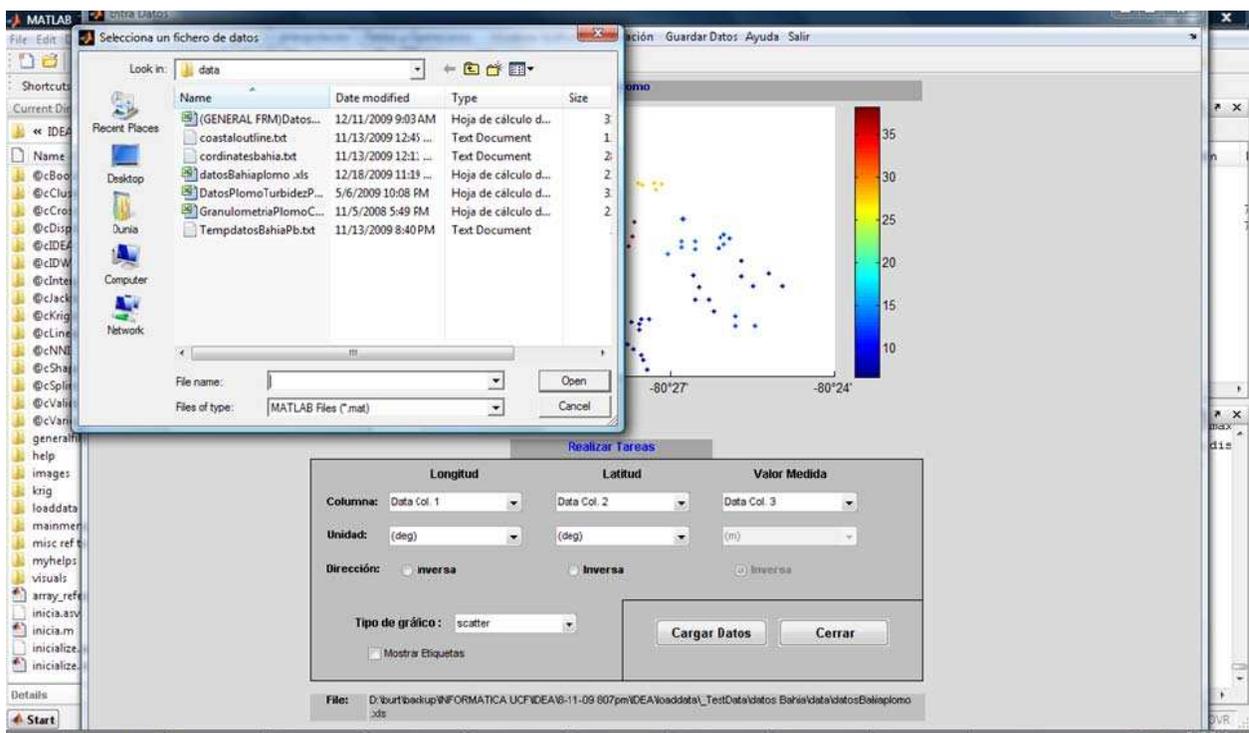
Vista de implementación:

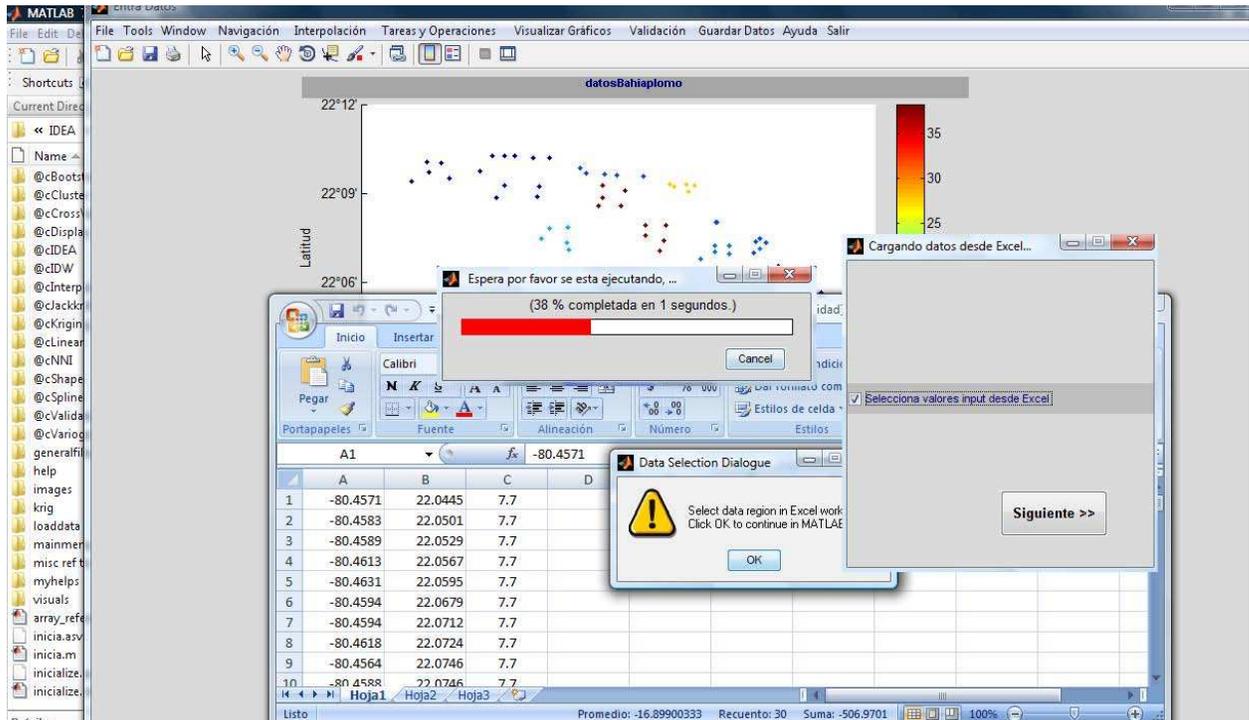
Comprende los componentes y los archivos que un sistema utiliza para ensamblar y hacer disponible el sistema físico. Se ocupa principalmente de la gestión de configuraciones de las distintas versiones del sistema. Los aspectos estáticos se capturan con los diagramas de componentes y los aspectos dinámicos con los diagramas de interacción, estados y actividades.

Anexo A

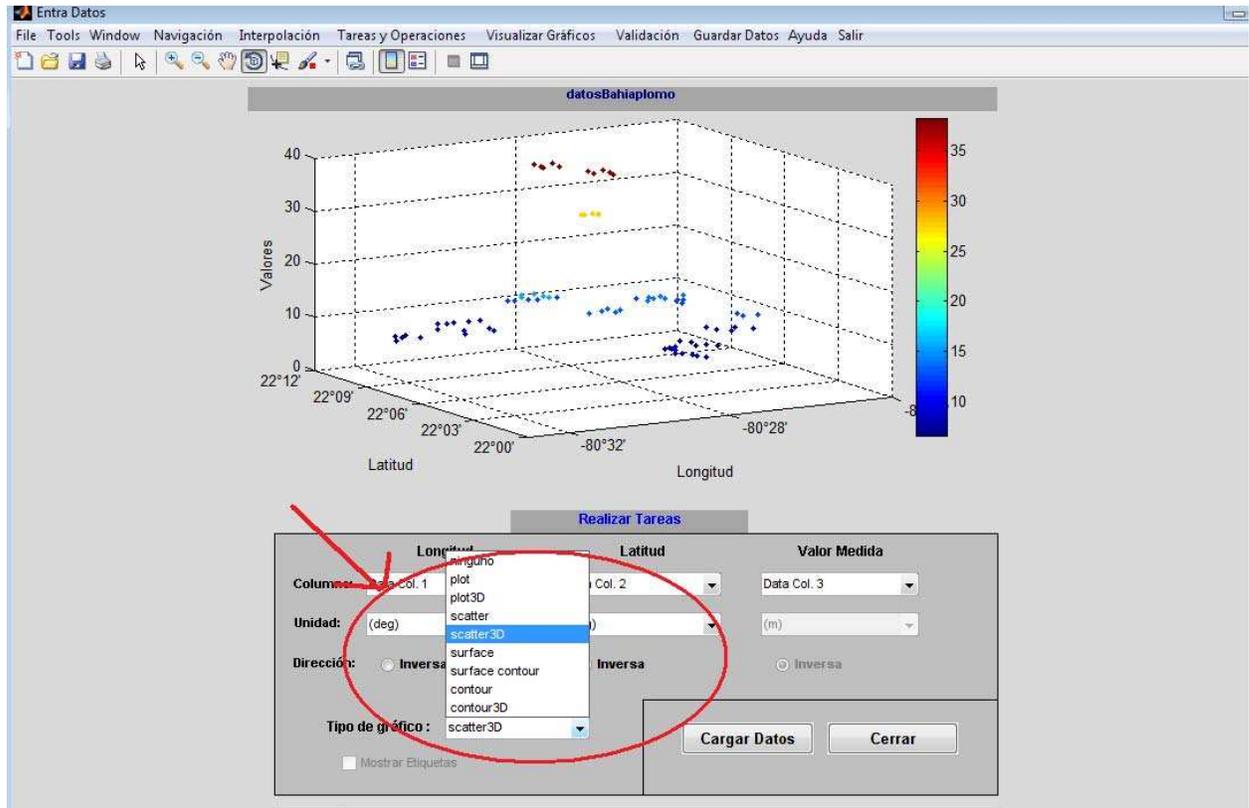
Prototipos de los Casos de Uso del sistema

A.1. Caso de Uso: Entrar datos espaciales

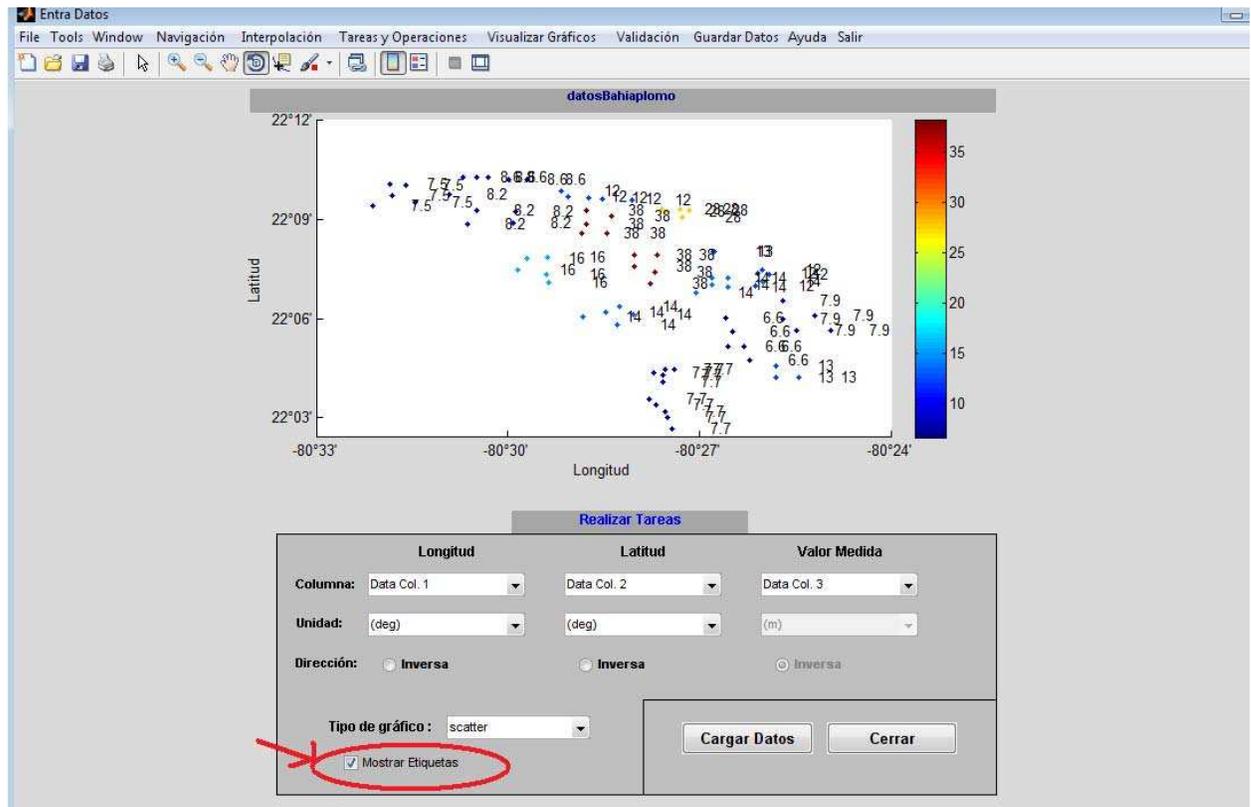




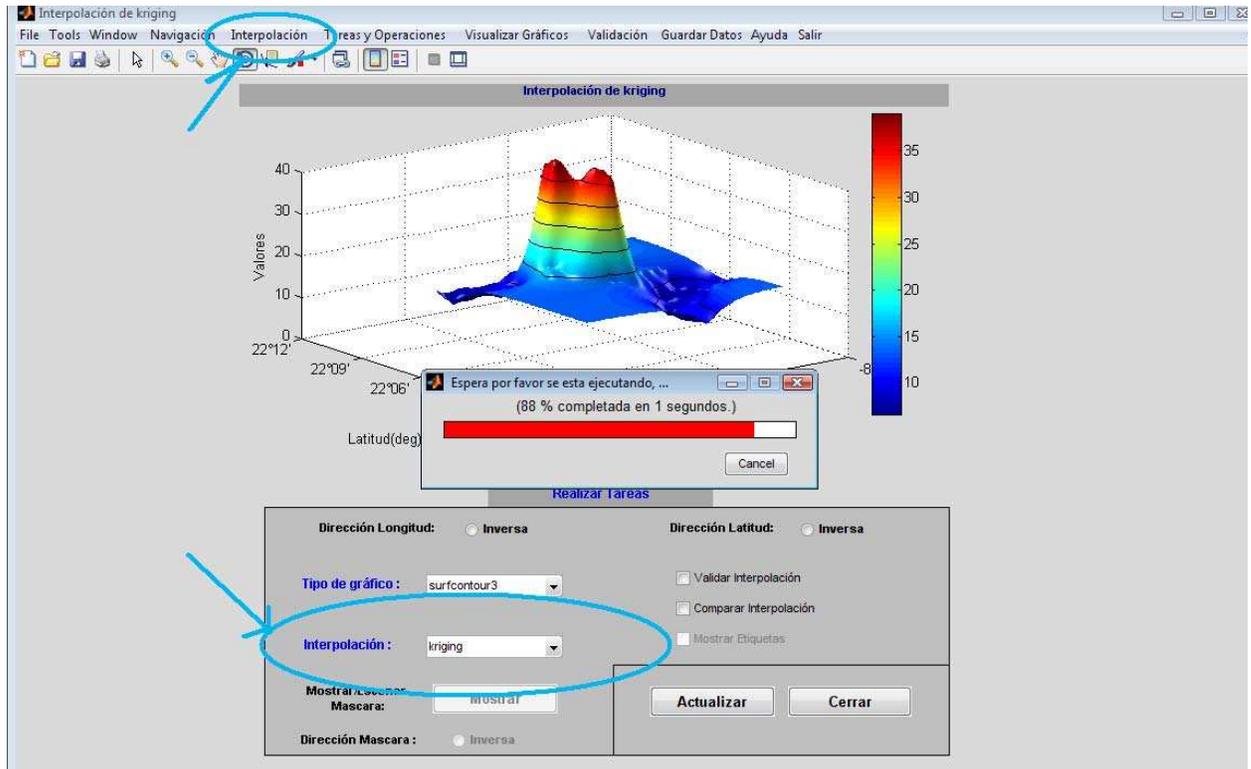
A.2. Caso de Uso: Representar gráficamente los datos crudos



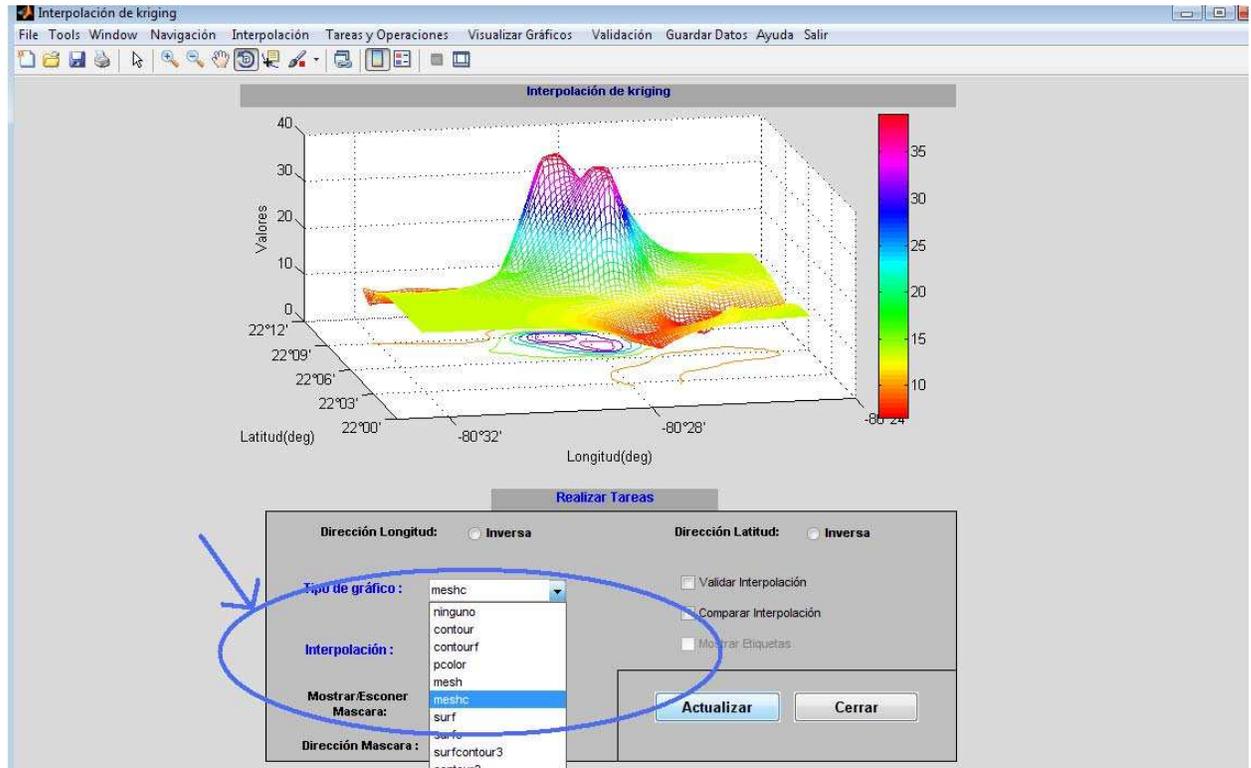
A.3. Caso de Uso: Representar gráficamente las etiquetas de los valores de los datos crudos



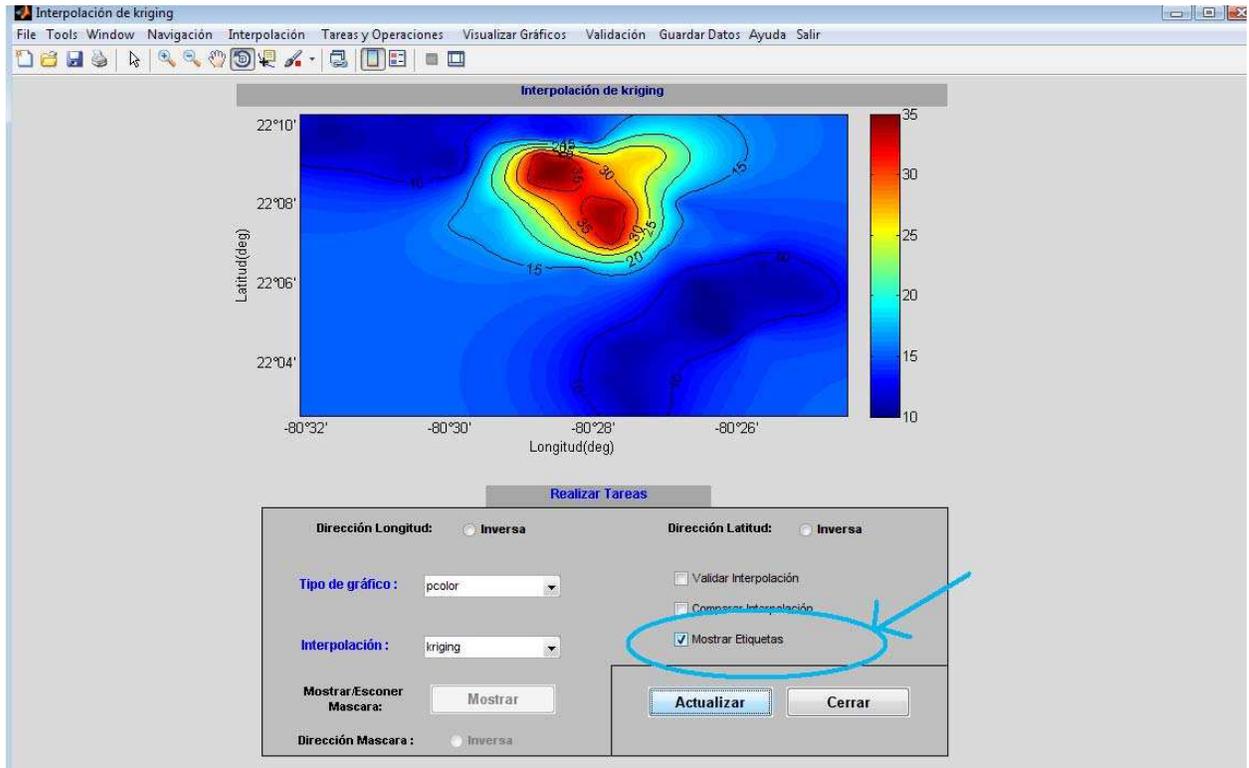
A.4. Caso de Uso: Interpolación de datos



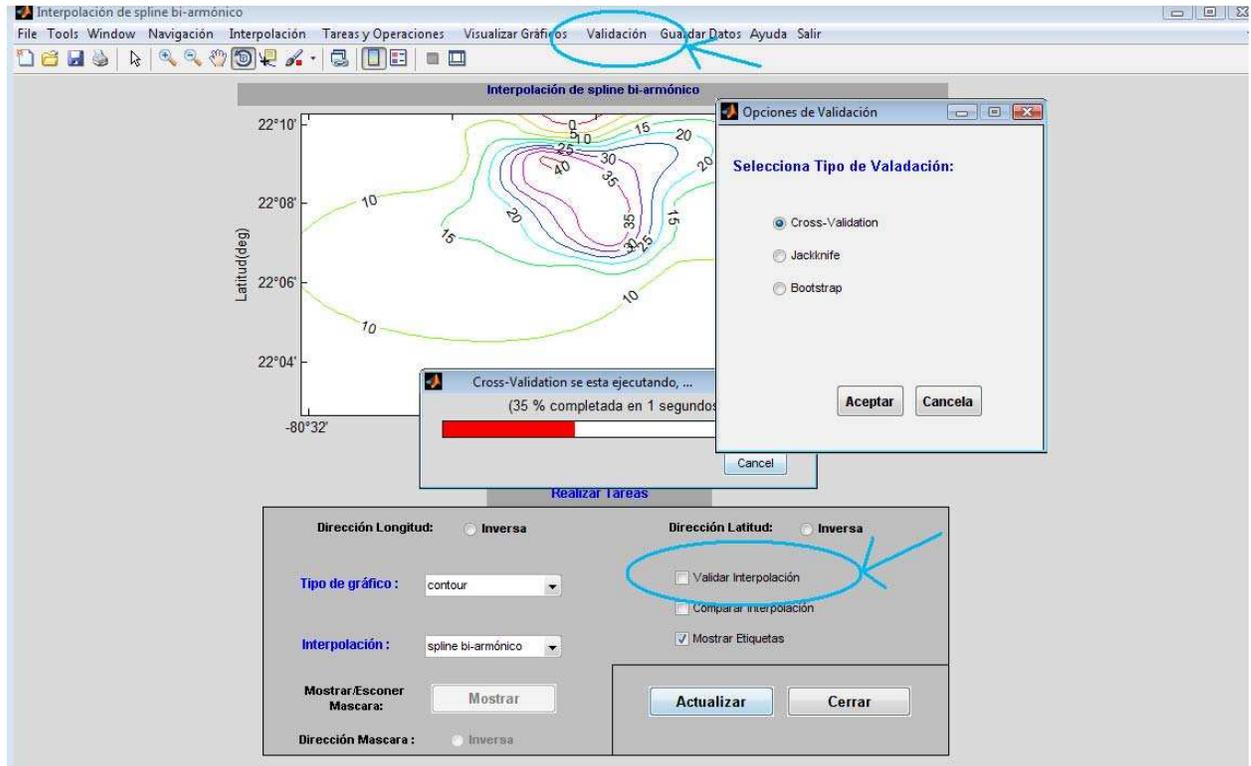
A.5. Caso de Uso: Representar gráficamente los datos interpolados



A.6. Caso de Uso: Representar gráficamente las etiquetas de los valores de los datos interpolados



A.7. Caso de Uso: Validar los datos interpolados



A.8. Caso de Uso: Generar reporte estadístico general

The screenshot shows the 'Validación de Cross-Validación' dialog box in a GIS application. The dialog contains a table with the following data:

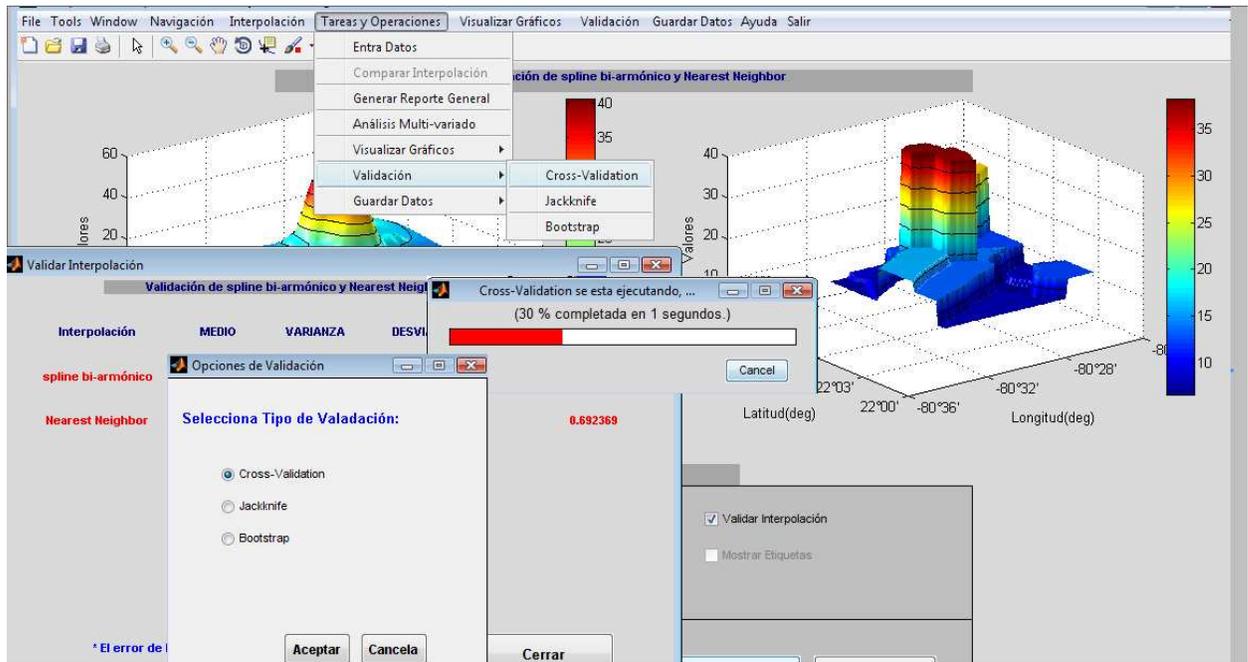
Interpolación	MEDIO	VARIANZA	DESVIACIÓN ESTANDAR	RMSE*
kriging	0.0157143	9.78247	3.12769	3.10812
spline cúbico	0.25682	8.22941	2.8687	2.65554
spline bi-armónico	0.0306738	5.75326	2.39859	2.38375
Inverse Distance Weig				3.85545
Nearest Neighbor				0.692369
Lineal				2.91402

Below the table, there is a note: '* El error de la'. At the bottom of the dialog, there are buttons for 'Mostrar', 'Actualizar', 'Cerrar', and 'Dirección Mascara: Inversa'.

Overlaid on the dialog is a 'Preparation de datos' dialog box with a question mark icon and the text: '¿Esta seguro que quiere termina el proceso?' with 'Yes' and 'No' buttons.

Another dialog box titled 'JACKKNIFE se esta ejecutando, ...' is also visible, showing a progress bar at 50% completion and a 'Cancel' button.

A.9. Caso de Uso: Generar reporte estadístico comparativo



A.10. Caso de Uso: Consultar ayuda

The screenshot displays the 'Ayuda y Referencia de IDEA' (IDEA Help and Reference) window. The main content area is titled '(Getting started) Trabajando con IDEA' and includes the following text:

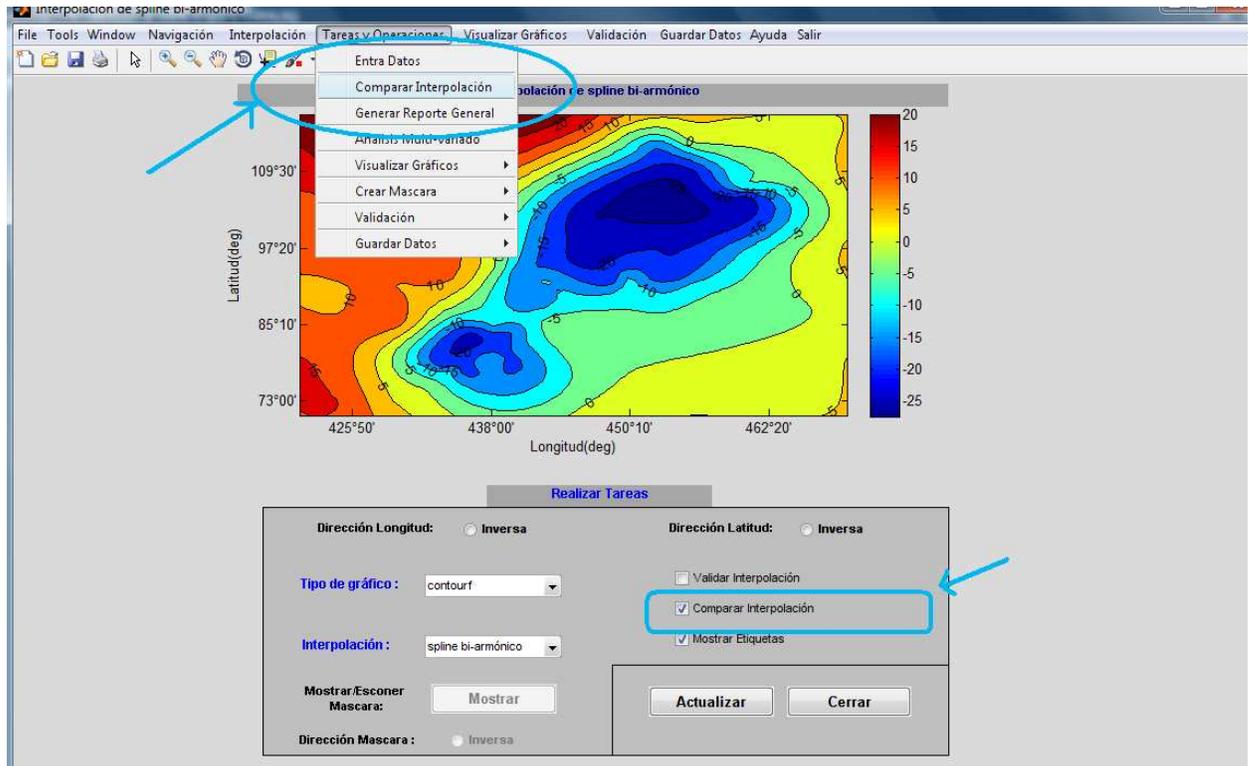
Rotar Gráfico
 En la barra de herramientas (TOOLBAR) se selecciona la herramienta de rotación, luego se hace clic en cualquier parte del eje (axis) del gráfico y con el mouse presionado hacer rotar la imagen. Ver la siguiente figura

The window also features a sidebar with a table of contents and a 3D surface plot. The sidebar lists the following sections:

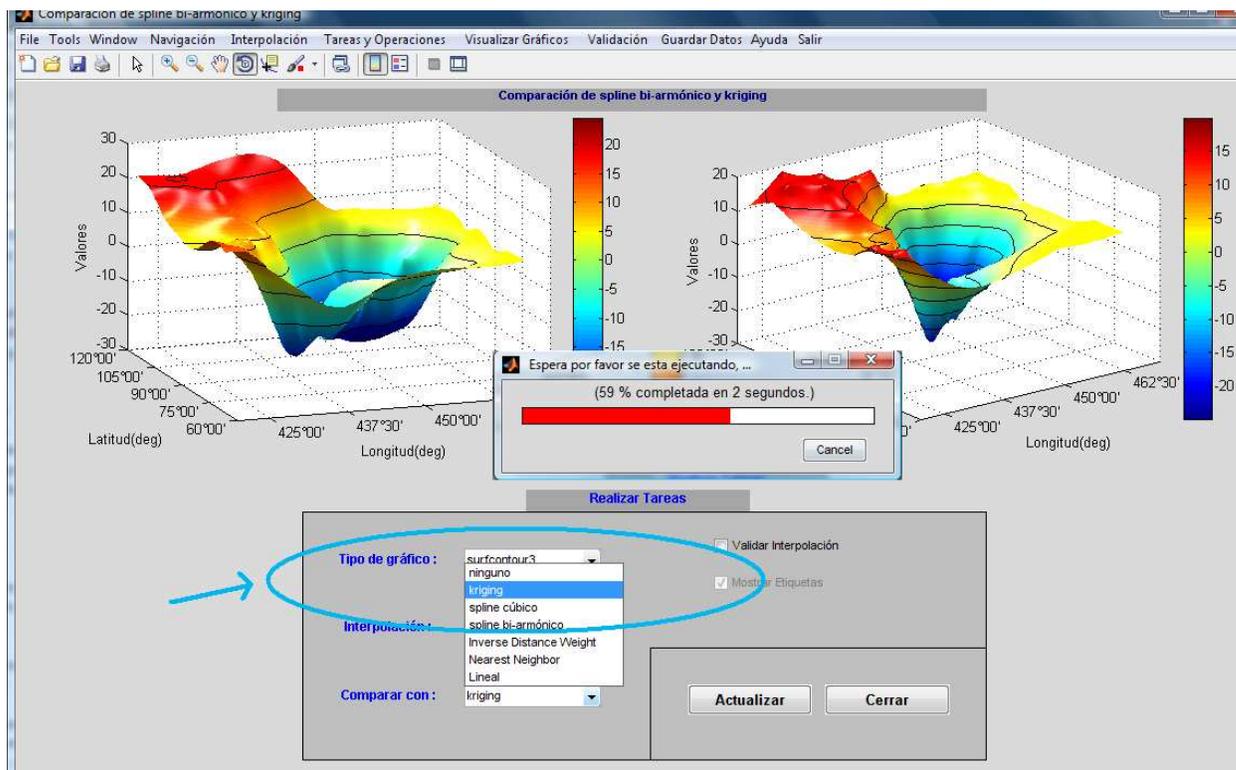
- **Interpolación**
 - ¿Qué es ID
 - Requerim
 - Instalaci
 - Iniciando
 - Utilizando
- **El ambient**
 - ¿Qué es M
 - Iniciando
 - El espaci
- **Preparaci**
 - Formato d
 - Preparanc
 - Preparanc
 - Preparanc
 - Preparanc
- **Trabajando en IDEA**
 - Entrar Datos
 - Visualizar Datos Crudos
 - Interpolar Datos
 - Rotar Gráfico
 - Comparar Interpoladores
 - Validar Interpoladores
 - Variograma
 - Guardar Datos

The 3D surface plot shows a topographic map with a color scale ranging from -25 to 20. The plot is displayed in a window titled 'Ayuda y Referencia de IDEA' and includes a toolbar with various icons for navigation and manipulation.

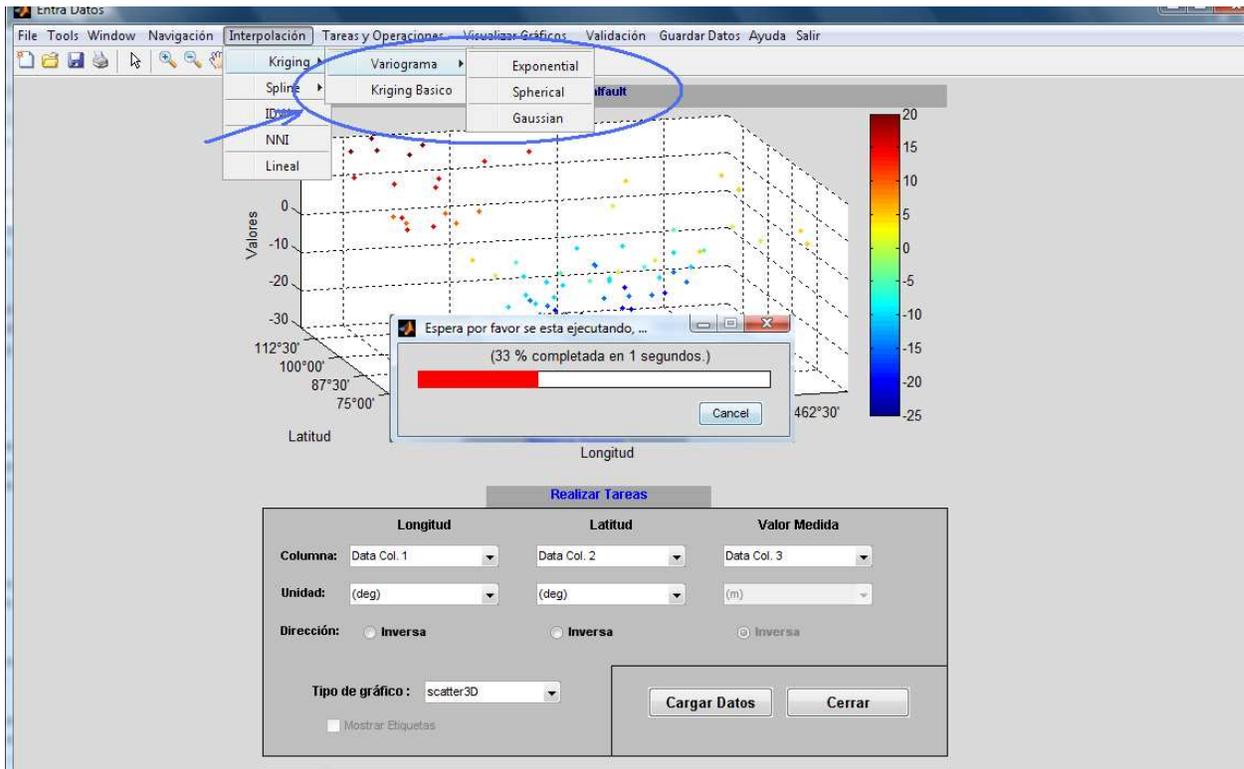
A.11. Caso de Uso: Comparar métodos de interpolación



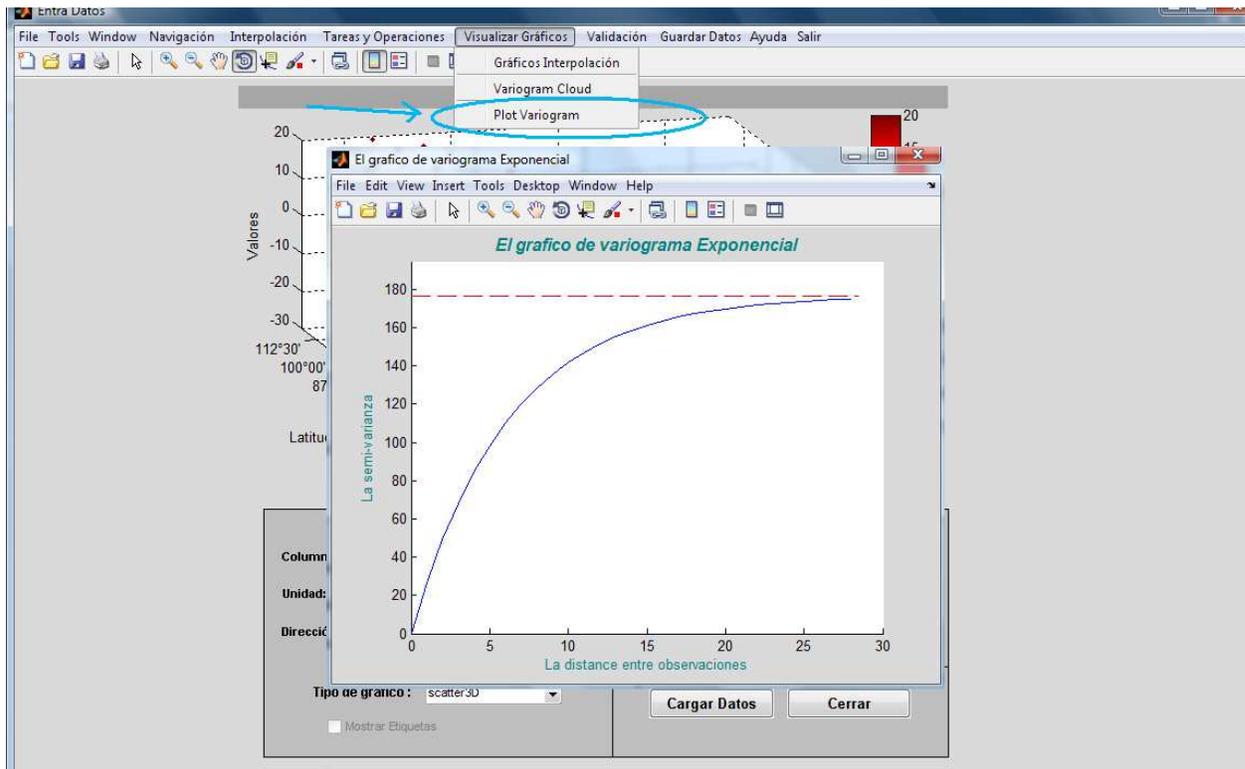
A.12. Caso de Uso: Representar gráficamente la comparación de interpoladores



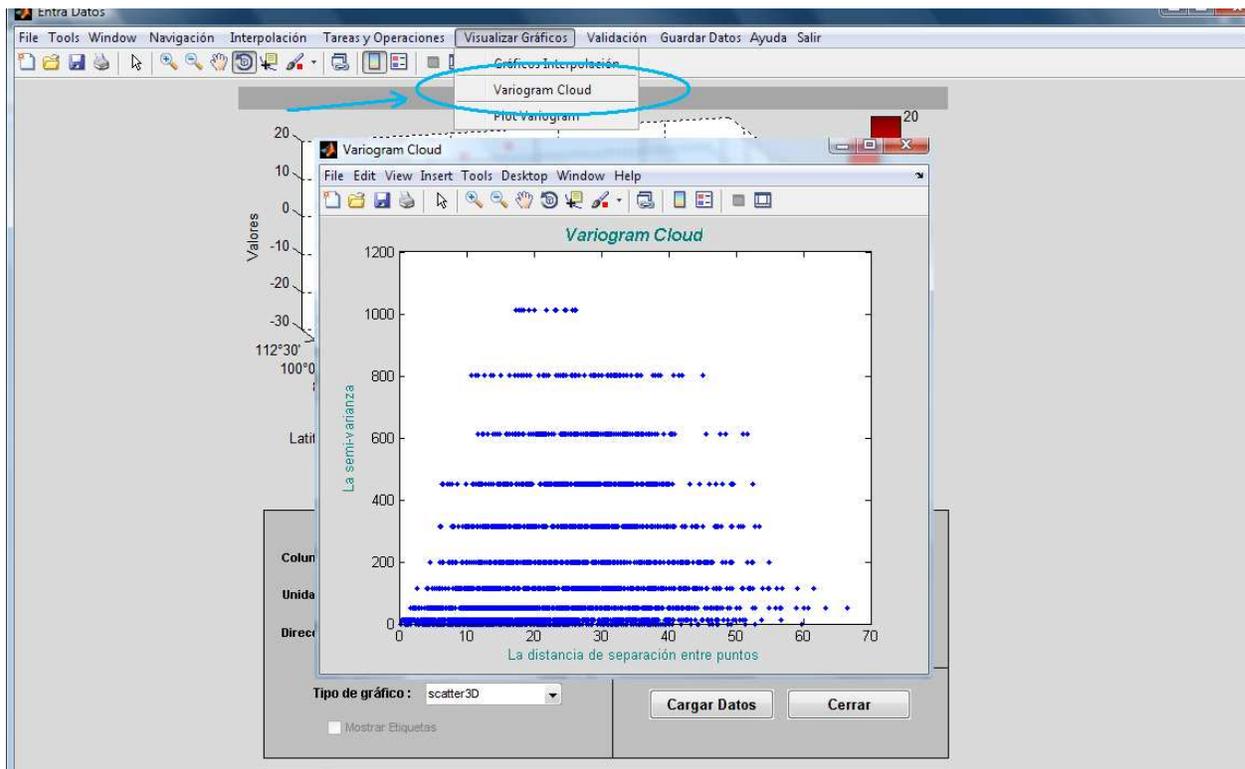
A.13. Caso de Uso: Crear modelo de variograma



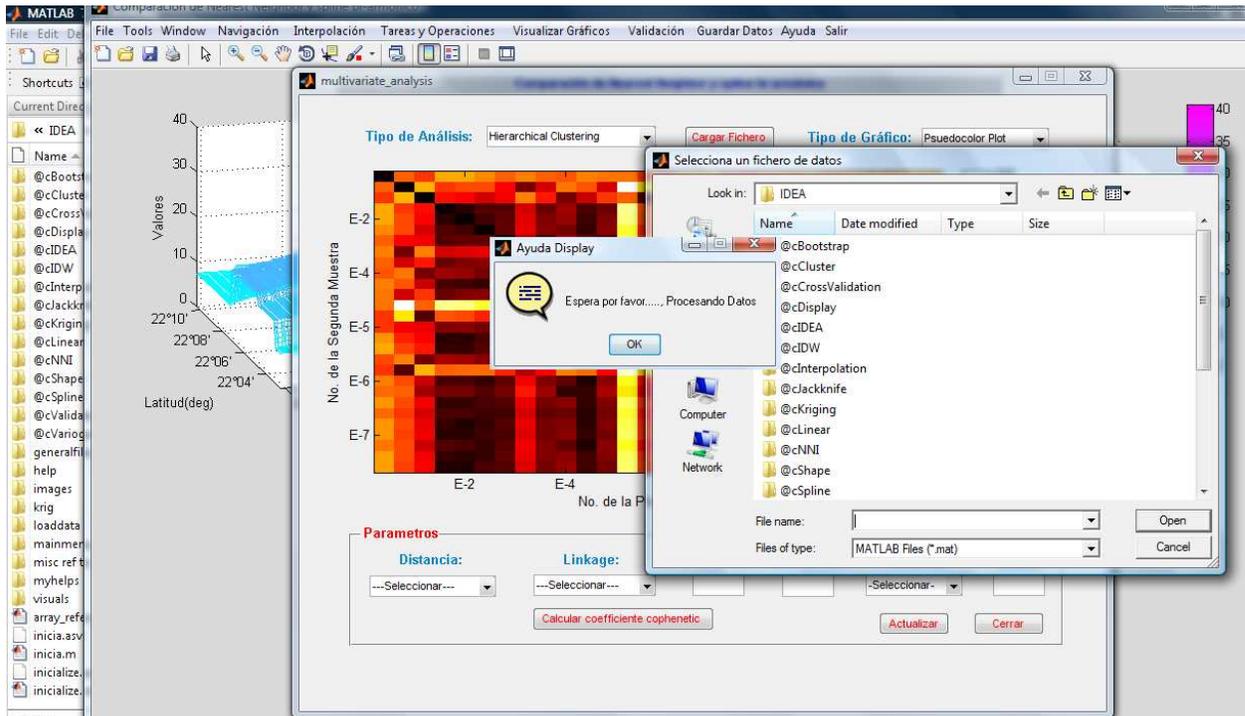
A.14. Caso de Uso: Representar gráficamente el variograma

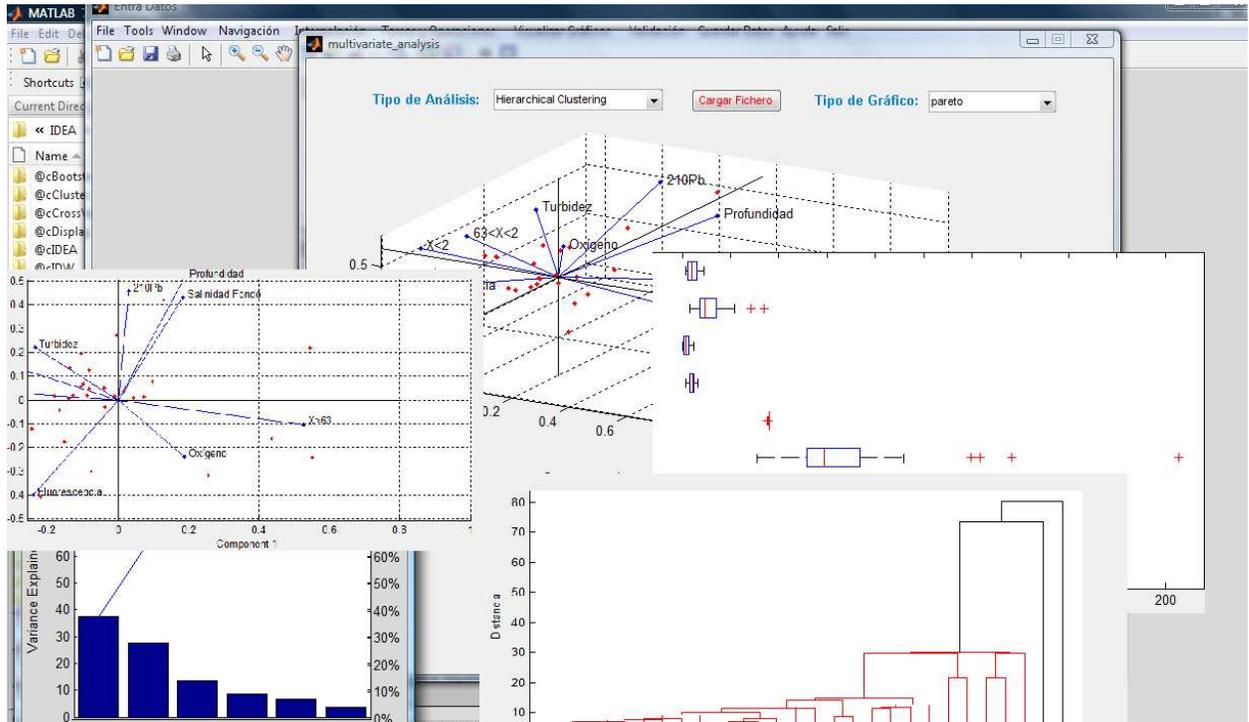


A.15. Caso de Uso: Crear nubes de variograma

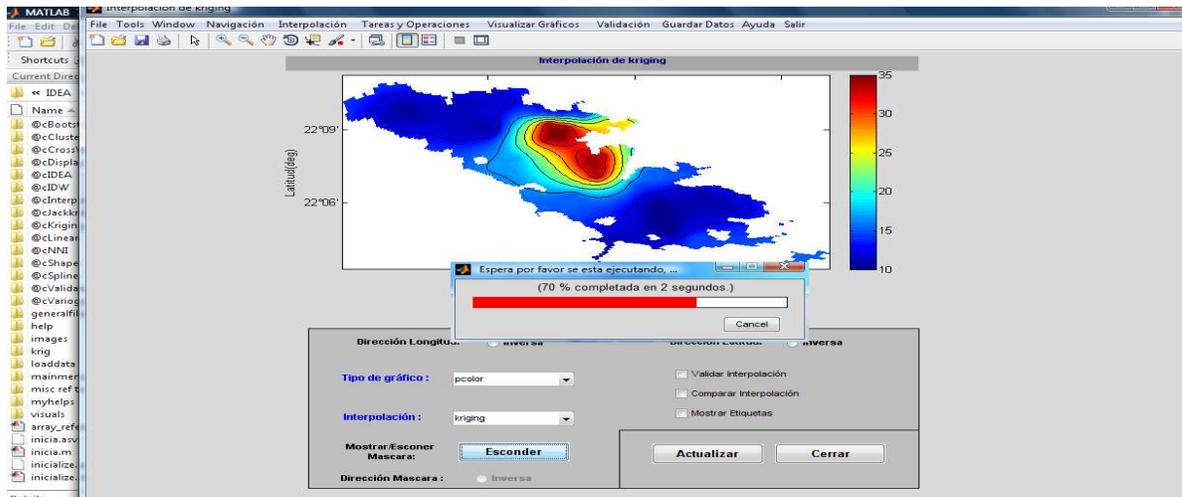
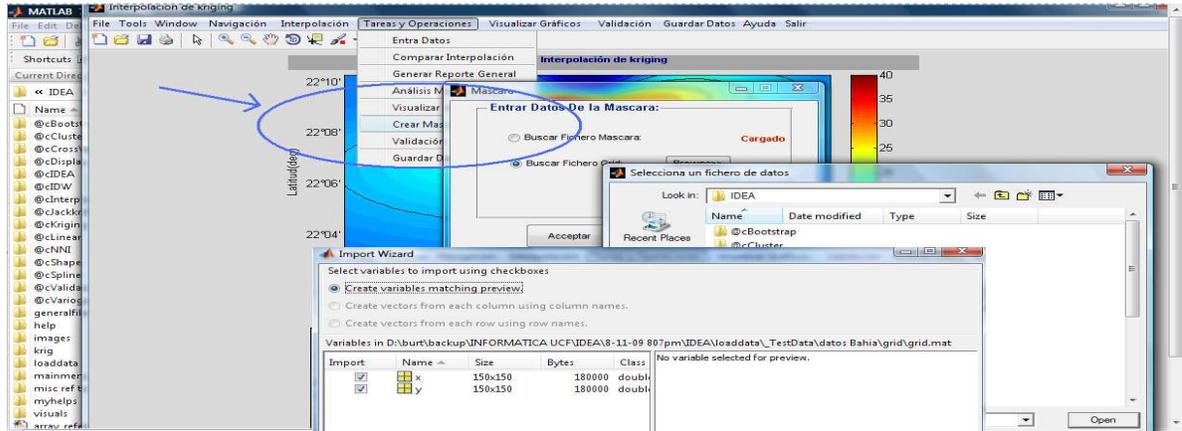


A.16. Caso de Uso: Realiza proceso de clúster

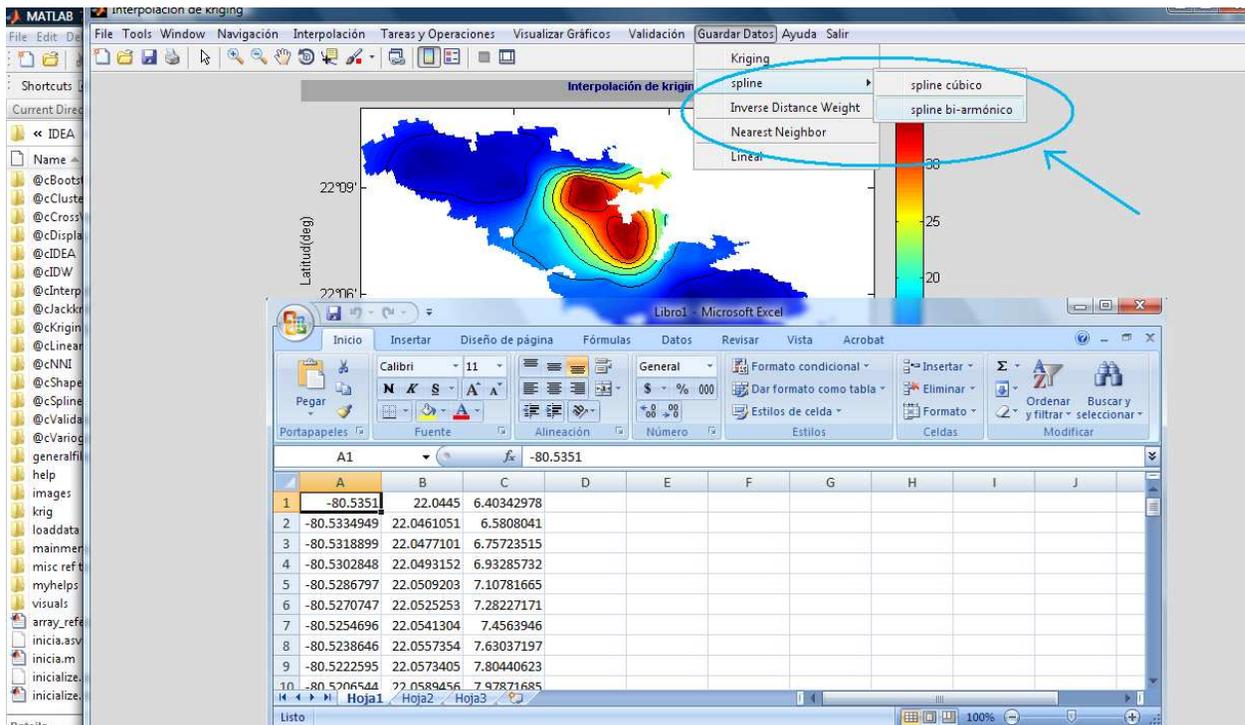




A.17. Caso de Uso: Crear máscara



A.18. Caso de Uso: Exportar datos interpolados



The screenshot shows the MATLAB interface for Kriging interpolation. The main window displays a map of a region with a color-coded interpolation result. A context menu is open over the map, showing options: Kriging, spline, Inverse Distance Weight, Nearest Neighbor, and Lineal. The 'spline' option is expanded, showing 'spline cúbico' and 'spline bi-armónico'. A blue circle highlights these two options, with an arrow pointing to 'spline bi-armónico'. Below the map, a Microsoft Excel window is open, displaying a table of data with columns A, B, and C, and rows 1 through 10. The data in column A is highlighted in yellow.

	A	B	C	D	E	F	G	H	I	J
1	-80.5351	22.0445	6.40342978							
2	-80.5334949	22.0461051	6.5808041							
3	-80.5318899	22.0477101	6.75723515							
4	-80.5302848	22.0493152	6.93285732							
5	-80.5286797	22.0509203	7.10781665							
6	-80.5270747	22.0525253	7.28227171							
7	-80.5254696	22.0541304	7.4563946							
8	-80.5238646	22.0557354	7.63037197							
9	-80.5222595	22.0573405	7.80440623							
10	-80.5206544	22.0589456	7.97871685							

Anexo B

B.1. Mex Executables (archivos ejecutables de MATLAB)

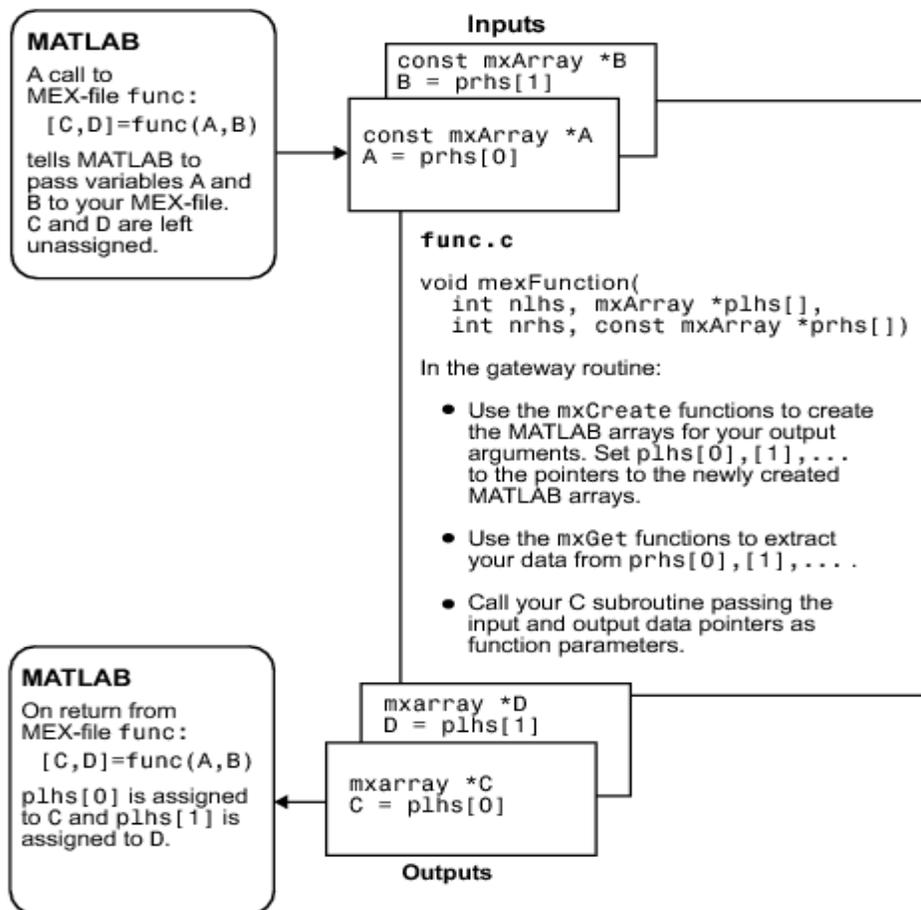


Figura B.1: El ciclo del proceso MEX en C/C++