

Especialización en Sistemas de Información Geográfica

Plugin en QGis para realizar predicciones espaciales mediante metodología INLA con enfoque SPDE

10 de octubre de 2022

Universidad Distrital Francisco José de Caldas

Vanessa Moyano 20212094019

Índice

1. Generalidades	5
1.1. Introducción	5
1.2. Planteamiento del problema	6
1.2.1. Antecedentes del problema	7
1.3. Justificación	8
1.4. Objetivos	9
1.4.1. Objetivo general	9
1.4.2. Objetivos específicos	9
2. Marcos de referencia	10
2.1. Marco conceptual	10
2.1.1. Dependencia y predicciones espaciales	10
2.1.2. Kriging	10
2.1.3. Probabilidad bayesiana	11
2.1.4. INLA	11
2.1.5. Plugin	11
2.1.6. Automatización de procesos	11
2.1.7. Usabilidad	12
2.1.8. Confiabilidad	12
2.1.9. Arquitectura de Software	12
2.1.10. Diseño de software	12
2.1.11. Calidad	13
2.2. Marco teórico	13
2.2.1. Regresión lineal	13
2.2.2. Campos Gaussianos y Campos Aleatorios de Markov Gaussianos	13
2.2.3. Aproximación de Laplace anidada integrada (INLA)	14
2.2.4. INLA con enfoque SPDE	15
3. Metodología	15
3.1. Método	15
3.2. Técnicas e instrumentos	16
4. Desarrollo de la propuesta	18
4.1. requerimientos del sistema	18
4.1.1. requerimiento funcional	18

4.1.2.	Requerimientos no funcionales	18
4.2.	Diseño de la arquitectura	18
4.2.1.	Diagrama de casos de uso:	18
4.2.2.	Diagrama de secuencia:	19
4.2.3.	Diagrama de clases:	20
4.2.4.	Diagrama de paquetes de clases:	21
4.2.5.	Diagrama de componentes:	21
4.2.6.	Diagrama de despliegue:	22
4.3.	Desarrollo	22
4.3.1.	Herramientas a utilizar	22
4.3.2.	Diseño de la interfaz gráfica	23
4.3.3.	Desarrollo del código lógico	24
4.4.	Guía de instalación	25
4.5.	Resultados	26
4.6.	Pruebas	28
4.6.1.	Prueba de rendimiento	29
4.6.2.	Usabilidad	30
5.	Conclusiones y recomendaciones	30
5.1.	Conclusiones	30
5.2.	Recomendaciones	31
	Referencias	32

Índice de figuras

1.	Metodología del proyecto	16
2.	Diagrama de casos de uso	19
3.	Diagrama de secuencia	20
4.	Diagrama de clases	20
5.	Diagrama de paquete de clases	21
6.	Diagrama de componentes	21
7.	Diagrama de despliegue	22
8.	Previsualización	23
9.	Maqueta del complemento	24
10.	Resultado espacial del modelamiento	27

11. Resultado estadístico del modelamiento 28

1. Generalidades

1.1. Introducción

Sin lugar a duda, los avances que ha tenido el desarrollo de complementos para herramientas bajo la filosofía de software libre han sido impulsados por las necesidades de crear de manera sencilla nuevas funcionalidades en programas comúnmente utilizados, automatizar tareas, promover avances científicos y aportar a la correcta toma de decisiones en campos como las ciencias de la tierra y el ordenamiento del territorio; ofreciendo instrumentos para manejar adecuadamente las diversas realidades que se presentan (Casanova, 2012). Conocer el potencial de las técnicas de análisis espacial mediante la interacción con Sistemas de Información Geográfica permite aprovechar diferentes mecanismos de reconocimiento de patrones complejos que pueden ser explotados al máximo con enfoques estadísticos como las predicciones espaciales, lo que permitiría robustecer los enfoques de análisis para investigación o toma de decisiones.

Las diferentes empresas y/o profesionales interesados en las ciencias de la tierra comúnmente se encuentran generando desarrollos y múltiples estudios ya sea para enfoques investigativos o administrativos y es natural pensar que no todos los interesados en estas ciencias están capacitados para resolver procesos estadísticos complejos en programas especializados o tienen equipos robustos para su procesamiento por lo que contar con diferentes herramientas integradas en un mismo programa sería de gran utilidad y facilidad para los interesados en el desarrollo de estas actividades analíticas (Niño Peña y Cáceres Castellanos, 2014).

Una posible estrategia para esta problemática es equipar un sistema de información geográfica de software libre como QGIS con un complemento de predicciones espaciales el cual fortalezca las funciones principales del programa implementando una metodología que permita reducir los

costos computacionales y resolver las dificultades que en ese sentido se tienen con los métodos tradicionales cuando se cuenta con bastantes datos, con el fin de que cualquier persona no necesariamente experta en temas estadísticos pueda obtener soluciones rápidas y funcionales para la consecución de resultados y toma de decisiones las cuales puedan aportar en la explicación o interpretación de los diferentes enfoques espaciales que se estén llevando para cualquier proyecto.

1.2. Planteamiento del problema

Las necesidades de las diferentes organizaciones o interesados para cumplir sus metas han planteado retos en cuanto al análisis de información (Niño Peña y Cáceres Castellanos, 2014), la automatización de tareas de geoprocésamiento es recomendable para todos los profesionales o técnicos en sistemas de información geográfica que necesiten realizar procesos y análisis espaciales; especialmente cuando se encuentran involucrados en proyectos con un volumen considerable de datos que requieran la implementación de diferentes herramientas de procesamiento (Morales, 2015) permitiendo que se pueden agrupar procesos disminuyendo los pasos que debe ejecutar un usuario para llegar a determinado producto (Valencia Cedeño, García Jimenez, Ascencio Arias, y cols., 2016).

En el campo de las geociencias, son comunes las variables espacialmente distribuidas (Mejía Rivera, Betancur Vargas, y Londoño Ciro, 2007) y el análisis espacial permite utilizar potentes técnicas de modelado para realizar predicciones y conocer mejor el mundo que nos rodea (*Análisis y ciencia de datos | Identificar patrones y realizar predicciones*, 2022), sin embargo, los métodos estadísticos más frecuentemente usados para la realización de estos análisis espaciales se caracterizan por costosas operaciones que repercuten en altos tiempos de ejecución, por lo que minimizar

el tiempo de ejecución de las aplicaciones para interpolación o predicción espacial puede ser un objetivo de alta prioridad (Fernández Díaz, Espinosa Ramirez, y García-Jacas, 2016).

1.2.1. Antecedentes del problema

A través de la historia se ha recurrido a diferentes técnicas que permitan reducir tiempos y mejorar la calidad de los procesos; pero es desde principios del siglo XX donde se comienza a implementar la automatización compleja (Agudelo, Tano, y Vargas, s.f.). En el campo de las ciencias de la tierra y el desarrollo de la estadística espacial se ha tenido un fuerte crecimiento en las últimas décadas ligado a la aparición y evolución de los SIG (Olaya, s.f.) que se ha visto fuertemente impulsada por la implementación de métodos numéricos y procesos computarizados los cuales han masificado la producción de información multiplicando exponencialmente la creación de datos geográficos (Siabato, s.f.).

Las predicciones espaciales han sido una herramienta para diferentes campos de la ciencia, en muchos casos se pueden agrupar de manera general en modelos probabilísticos basados en teoría bayesiana, modelos estadísticos basados en análisis de correlación, estadística espacial, y modelos de aprendizaje automático los cuales han tomado importancia en los últimos años (García y Castillo, 2019). En los trabajos anteriores de Kriging, se utilizan interpolaciones clásicas como una combinación lineal de las n variables aleatorias distribuidas en todo el dominio (Romero Rios y Castaño Salazar, 2021), lo que conlleva a problemas computacionales cuando n es demasiado grande dentro de la teoría de los Campos Gaussianos (GF, siglas en ingles). Por lo tanto, nace el planteamiento propuesto por (Lindgren, Rue, y Lindström, 2011), en el cual, se plantea una relación entre un GF y un Campo Aleatorio de Markov Gaussiano (GMRF, siglas en ingles) mediante el enfoque de Ecuaciones Diferenciales Parciales Estocásticas (SPDE, siglas en ingles). Un

ejemplo de la aplicación de este enfoque lo realiza (Gómez-Rubio, 2020), en su libro Inferencia bayesiana con INLA, en el que compara dos modelaciones básicas de kriging y SPDE con el fin de estimar concentraciones de metales pesados en cualquier punto de una región de estudio. Otro ejemplo en donde se incorpora el enfoque SPDE para predicción de datos espaciales continuos se encuentra en el libro de (Blangiardo y Cameletti, 2015) Modelos bayesianos espaciales y espacio-temporales con R-INLA, donde trata el tema de los modelos espaciales con enfoque SPDE, de manera similar (Moraga, 2019) utiliza los datos geoespaciales en el campo de la salud y las políticas publicas describiendo métodos geoestadísticos y espacio-temporales para analizar datos de salud mediante un modelado espacial utilizando R-INLA en R.

1.3. Justificación

Es natural pensar que no todos los encargados de realizar análisis espaciales cuentan con amplios conocimientos estadísticos y manejo de programas especializados lo que lleva a incurrir en otros gastos como consolidar equipos de trabajo con diferentes profesionales que cuenten con las capacidades estadísticas y el manejo de programas especializados y por otro lado en los perfiles que analizaran la información procesada en pro de sus objetivos, así mismo se debe pensar en adquirir equipos de cómputo capaces de procesar adecuadamente la cantidad de información con la que se pretenda realizar los diferentes proyectos.

En el campo de la estadística espacial es importante visualizar los beneficios de alternativas estadísticas novedosas para realizar predicciones espaciales como la metodología de Aproximación de Laplace anidada integrada (INLA) la cual es relativamente reciente y una de las motivaciones principales que tiene es resolver la dificultad computacional de otros métodos avanzados de

muestreo que generalmente están basados en estadística frecuentista o inferencia Bayeciana como Markov Chain Montecarlo (MCMC) (*R-INLA Project*, 2022) ya que estos algoritmos necesitan altos tiempos de ejecución y conocimiento en temas geo-estadísticos para la evaluación de supuestos, mientras que las aproximaciones por INLA proporcionan estimaciones más precisas hasta en segundos y minutos (M. R. Ruiz, 2020). También es importante resaltar que algunos de los últimos estudios han demostrado que ha tenido mejores predicciones que Krigin o Cokrigin evitando validación de supuestos (Romero Rios y Castaño Salazar, 2021).

La automatización de los procesos aprovechando los lenguajes de programación y diferentes algoritmos hace que las tareas se realicen mucho más rápido (Arriola, 2021) es por esto que se hace mención de las ventajas de los plugins o complementos para software que permiten equipar fácilmente a los programas y aplicaciones con nuevas funcionalidades (*¿Qué es un plugin y para qué se usa?*, 2020).

1.4. Objetivos

1.4.1. Objetivo general

- Desarrollar y evaluar el funcionamiento de un complemento para el software QGIS que realice predicciones espaciales por metodología R-INLA con enfoque SPDE.

1.4.2. Objetivos específicos

- Diseñar la arquitectura de software que responda a los requerimientos funcionales y no funcionales del plugin.
- Desarrollar el complemento para QGIS combinando lenguajes de programación como python

y R que permita implementar la metodología R-INLA.

- Realizar pruebas para evaluar la funcionalidad, usabilidad y rendimiento del complemento.

2. Marcos de referencia

2.1. Marco conceptual

2.1.1. Dependencia y predicciones espaciales

La información obtenida para uso de la ciencia regional posee características que provienen de su ubicación geográfica, esta información posee características que constituyen los denominados efectos espaciales (Bohórquez y Ceballos, 2008) uno de ellos es la dependencia espacial que consiste en la falta de independencia que se produce a menudo entre las observaciones de datos espaciales cercanos, en línea con los contenidos formulados por (Tobler, 1969) en su “primera ley de la geografía” en la que se establece en la cual “todo esta relacionado con todo lo demás, pero las cosas cercanas están más relacionadas que las cosas lejanas”, la dependencia espacial podría ser definida como la existencia de una relación funcional entre lo que ocurre en un punto determinado del espacio y lo que ocurre en otro lugar (Serrano y Valcarce, 2000), esta dependencia espacial es analizada y modelada a través de métodos matemáticos y estadísticos que permiten predecir o interpolar los valores de una variable en sitios donde no fueron muestreados.

2.1.2. Kriging

Es un conjunto de métodos de predicción espaciales basados en la estadística frecuentista en el que se realizan análisis de un espacio continuo que nos permite predecir o interpolar los valores

de una variable en lugares no fueron muestreados utilizando la información proporcionada por la muestra (Díaz y Cocunubo, 2021).

2.1.3. Probabilidad bayesiana

Está basada en la interpretación subjetiva de la probabilidad y tiene como punto central el Teorema de Bayes demuestra los beneficios obtenidos en las estimaciones basadas en conocimientos intrínsecos o información conocida previamente (Páez, 2011).

2.1.4. INLA

La aproximación de Laplace anidada integrada (INLA) es un método para la inferencia bayesiana aproximada. En los últimos años se ha consolidado como una alternativa a otros métodos tradicionales por su rapidez y facilidad de uso a través del paquete R-INLA en RStudio (*R-INLA Project*, 2022).

2.1.5. Plugin

Un plugin es aquella aplicación que, en un programa informático, añade una funcionalidad adicional o una nueva característica al software. En nuestro idioma, por lo tanto, puede nombrarse al plugin como un complemento (*¿Qué es un plugin y para qué se usa?*, 2020).

2.1.6. Automatización de procesos

La automatización es una importante potenciación de las virtudes de un proceso, obteniendo mejoras en cuanto a rendimiento, eficiencia y productividad para una organización o persona interesada. También denominada automatización de la infraestructura, consiste en el uso de sistemas

de software para crear instrucciones y procesos repetibles a fin de realizar las tareas con muy poca intervención humana, o sin ella (*¿Qué es la automatización?*, 2022).

2.1.7. Usabilidad

La Usabilidad es la medida de la calidad de la experiencia que tiene un usuario cuando interactúa con un producto o sistema. Esto se mide a través del estudio de la relación que se produce entre las herramientas y quienes las utilizan, para determinar la eficiencia en el uso de los diferentes elementos ofrecidos en las pantallas y la efectividad en el cumplimiento de las tareas que se pueden llevar a cabo a través de ellas (*¿Qué es la Usabilidad? | Guía Digital*, 2022).

2.1.8. Confiabilidad

Es la habilidad que tiene un sistema o componente de realizar sus funciones requeridas bajo condiciones específicas en periodos de tiempo determinados (*Confiabilidad en el Software*, 2016).

2.1.9. Arquitectura de Software

La Arquitectura de Software es la organización fundamental de un sistema enfocada en sus componentes, las relaciones entre ellos, el ambiente y los principios que orientan su diseño y evolución (Hilliard, 2000).

2.1.10. Diseño de software

Es la descripción de la arquitectura del software cómo va a estar compuesto y organizado, las interfaces entre dichos componentes, y los componentes a un nivel de detalle que permita su adecuada construcción (F. Ruiz, s.f.).

2.1.11. Calidad

La calidad del software es el conjunto de cualidades que lo caracterizan y que determinan su utilidad y existencia adecuada (García León y Beltrán Benavides, s.f.) cumpliendo los requerimientos del usuario.

2.2. Marco teórico

2.2.1. Regresión lineal

El modelo de regresión es un método estadístico que se utiliza ampliamente en modelos cuantitativos. La regresión lineal es un enfoque básico y estándar en el que se establece una relación lineal entre los valores de variables predictoras y un resultado como se ve en la ecuación (1). La regresión lineal Bayesiana univariada es un enfoque de Regresión lineal donde el análisis estadístico se realiza dentro del contexto de la inferencia Bayesiana (Statistics, 2021)

$$\mu_i = \beta_0 + Z_{s_i} \quad (1)$$

2.2.2. Campos Gaussianos y Campos Aleatorios de Markov Gaussianos

Los campos Gaussianos, GF (siglas en inglés), constituyen un componente importante de los modelos jerárquicos espaciales actuales (Banerjee, Carlin, Gelfand, y Banerjee, s.f.), como el caso de la aplicación de kriging, que satisface las condiciones de las funciones de correlación espacial y validación de supuestos. Los GF son uno de los pocos modelos multivariantes apropiado con una constante de normalización explícita y con buenas propiedades analíticas, sin embargo llevan consigo un problema de alto costo computacional, cuando n es demasiado grande, lo cual implica

la factorización de matrices de covarianza de orden $n \times n$ (Monsalve Graterol, 2013). Esto debido a que al momento de aplicar inferencia estadística generalmente es necesario evaluar la función de probabilidad o la distribución latente del campo gaussiano, para lo cual necesitamos hacer cálculos extensos con matrices. Para facilitar la realización del análisis cuando n es grande, se aplica el enfoque propuesto por (Lindgren y cols., 2011), en el cual un GF es reemplazado por un Campo Aleatorio de Markov Gaussiano (GMRF, siglas en inglés), la notación matemática que representa la distribución de un GMRF: $x \sim N(\mu, Q^{-1})$ donde μ es la media y Q matriz de precisión simétrica y definida positiva, Q (inversa de la matriz de covarianza) (Monsalve Graterol, 2013)

Una función de correlación muy popular es la función de correlación de Matern, que depende de un parámetro de escala $\kappa > 0$ y un parámetro de suavidad $\nu > 0$. Considerando dos ubicaciones s_i y s_j , la función de correlación de Matern estacionaria e isotrópica es:

$$Cor_M(X(s_i), X(s_j)) = \frac{2^{1-\nu}}{\Gamma(\nu)} (\kappa \|s_i - s_j\|)^\nu K_\nu(\kappa \|s_i - s_j\|) \quad (2)$$

donde $\| \cdot \|$ denota la distancia euclidiana y K_ν es la función de Bessel modificada de segundo orden. La función de covarianza de Matern es $\sigma_x Cor_M(X(s_i), X(s_j))$, donde σ_x es la varianza marginal del proceso.

2.2.3. Aproximación de Laplace anidada integrada (INLA)

La aproximación posterior conjunta puede ser generada usando tres enfoques diferentes: la aproximación Gaussiana, una aproximación completa de Laplace y una simplificada de Laplace. INLA es una versión simplificada, rápida de calcular y genera aproximaciones lo suficientemente precisas (Havard Rue, 2009) debido a que simplifica los costos computacionales reduciendo el

tiempo de ejecución al método tradicional MCMC y permite realizar análisis bayesianos de forma simplificada, al igual que ofrece criterios de comparación para la elección del mejor modelo (Romero Rios y Castaño Salazar, 2021) los autores desarrollaron un modelo computacional novedoso para la inferencia bayesiana en el que se enfocan en obtener aproximaciones a distribuciones posteriores univariadas, a su vez implementaron un paquete de R para facilitar el ajuste del modelo.

2.2.4. INLA con enfoque SPDE

La implementación de regresión bayesiana para datos espaciales estimada por INLA utilizando el enfoque de Ecuaciones Diferenciales Parciales Estocásticas (SPDE) para estimar la estructura de correlación espacial permite ajustar GF por una aproximación del GF a cualquier GMRF y en lugar de establecer una red de vecindarios para definir la estructura espacial se utiliza una malla construida por triangulación. La estimación del efecto aleatorio espacial se realiza a partir de una función de Mátern que se resuelve por SPDE. Luego, la estimación de este efecto espacial se proyecta utilizando la malla sobre los sitios de predicción. Este mecanismo hace que no sea necesario calcular grandes matrices de distancias para la estimación de la correlación espacial como sí ocurre con los otros métodos, lo que implica un mayor trabajo computacional (Giannini-Kurina, Suarez, Paccioretti, Macchiavelli, y Mónica, s.f.).

3. Metodología

3.1. Método

Para desarrollar este complemento es necesario ejecutar una serie de tareas que permitirán cumplir con el objetivo y las reglas de negocio que se planteen para el adecuado funcionamiento

del plugin, se ha seleccionado un enfoque ágil mediante una metodología que combine modelos como el incremental el cual permite realizar entregas más rápidas y ajustarse a posibles modelos cambiantes y el modelo orientado a la reutilización el cuál permite ahorrar trabajo dado el tiempo que se tiene para la ejecución del proyecto.

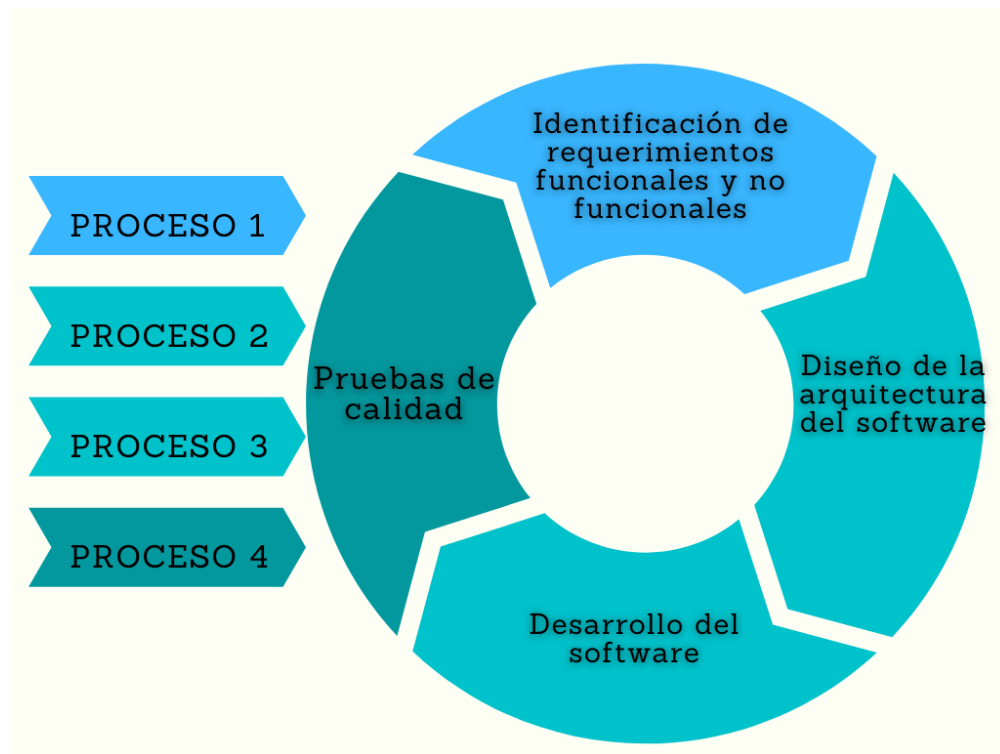


Figura 1: Metodología del proyecto

3.2. Técnicas e instrumentos

- **Proceso 1 - Ingeniería de requerimientos:** Parte de la Ingeniería de Software encargada de abordar la definición de los servicios que el sistema debe proporcionar y de establecer las restricciones operativas del mismo (Briseño, s.f.), es por esto que se considera como el primer proceso en el desarrollo del complemento para QGIS, debe ser un diálogo entre el usuario y el ingeniero de requerimientos para cumplir con cada una de las expectativas que se

tiene sobre el proyecto; el resultado de este proceso son los requerimientos funcionales y no funcionales que permiten establecer una guía de como debe ser el diseño del complemento.

- **Proceso 2 - Arquitectura de software:** es la parte de la Ingeniería de Software encargada de representar el sistema, incluyendo los componentes principales del mismo, las relaciones e interacciones que se presentan entre estos para alcanzar el objetivo del sistema, aportando el más alto nivel de comprensión y detalle sobre lo que se quiere lograr. (Reynoso, 2004) se pretende establecer una guía totalmente detallada para el adecuado cumplimiento del desarrollo, el resultado de este proceso son los diagramas de casos de uso, secuencia, clases y paquetes que permiten visualizar el comportamiento y como esta compuesto el sistema a desarrollar.
- **Proceso 3 - Desarrollo de software:** Para el desarrollo de complementos en QGis es necesario utilizar el lenguaje de programación python y para el caso de la metodología INLA está se encuentra implementada en un paquete de programación de R por lo que es necesario utilizar las diferentes ventajas que tiene python y así poder combinar los dos lenguajes de programación llevando a cabo el proceso de desarrollo en función de lo diseñado en la arquitectura de software, el resultado de este proceso es el código lógico del complemento y la interfaz gráfica que le permitirá al usuario realizar las predicciones espaciales.
- **Proceso 4 - Pruebas:** En este último proceso se pretende implementar pruebas de validación para el complemento mediante el método no probabilístico de muestreo por conveniencia el cual es útil cuando se tiene tiempo y recursos limitados (Etikan, Musa, Alkassim, y cols., 2016) con el fin de evaluar la usabilidad y recursos básicos necesarios para el funcionamiento, por lo que será compartido con mínimo tres personas considerando el nivel de comple-

idad que tiene el proyecto un número suficiente para la validación (*No me hagas pensar. Actualización - Anaya Multimedia, 2015*).

4. Desarrollo de la propuesta

4.1. requerimientos del sistema

4.1.1. requerimiento funcional

- **RF-1 Predicciones espaciales:** Plugin para Qgis que realice predicciones espaciales por la metodología R-INLA.

4.1.2. Requerimientos no funcionales

Producto - usabilidad	
RNF-1 Intuición	La interfaz debe ser intuitiva con el fin de reducir tiempos de entrenamiento soporte y prueba
RNF-2 Lógica	Debe garantizar la lógica de flujo de eventos asociado a cada uno de los elementos de la interfaz
Producto - disponibilidad	
RNF-3 Bajas fallas	Bajo nivel de incidencia a fallas mientras está en funcionamiento
Organizacional - entorno	
RNF-4 Acceso	El complemento debe ser accedido desde QGis desktop
Organizacional - operacional	
RNF-5 Idioma	El complemento debe ser implementado en idioma español

4.2. Diseño de la arquitectura

4.2.1. Diagrama de casos de uso:

Este diagrama permite observar la relación que tiene el actor con el sistema para llevar a cabo el caso de uso, para el caso de este complemento se tiene un solo actor ya que el sistema funcionará

de la misma manera para cualquier usuario.

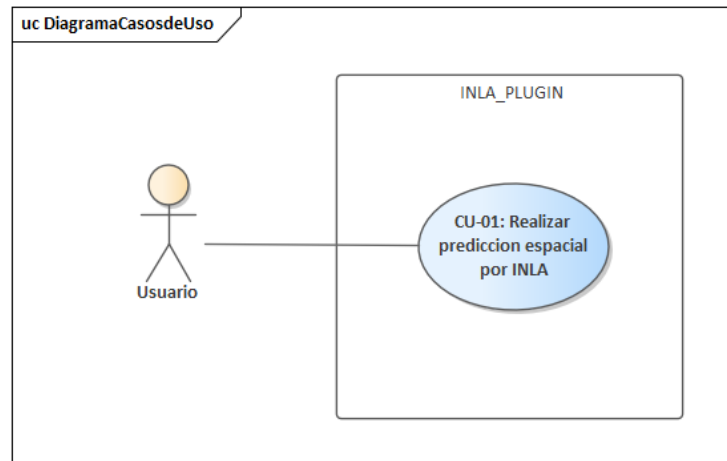


Figura 2: Diagrama de casos de uso

4.2.2. Diagrama de secuencia:

Este diagrama se centra en representar la interacción del sistema describiendo cómo y en qué orden los actores y objetos funcionan en conjunto para cumplir con el caso de uso; en este caso realizar las predicciones espaciales.

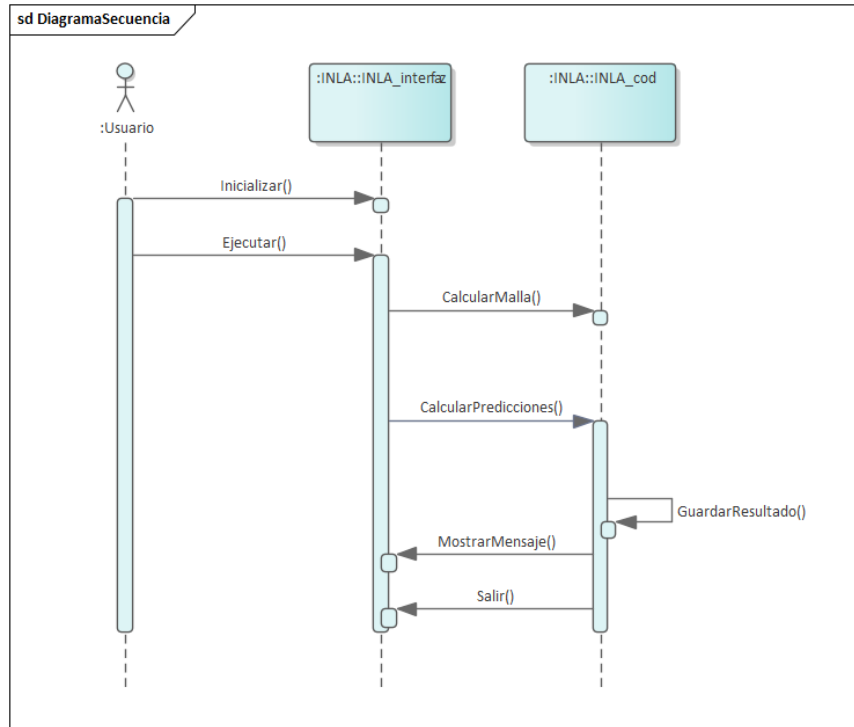


Figura 3: Diagrama de secuencia

4.2.3. Diagrama de clases:

Este diagrama permite ver la estructura del sistema con sus respectivas clases para su adecuado funcionamiento, se observa que el diagrama cuenta con dos clases que componen el plugin como son la interfaz gráfica de usuario y el código que tendrá la lógica del sistema.

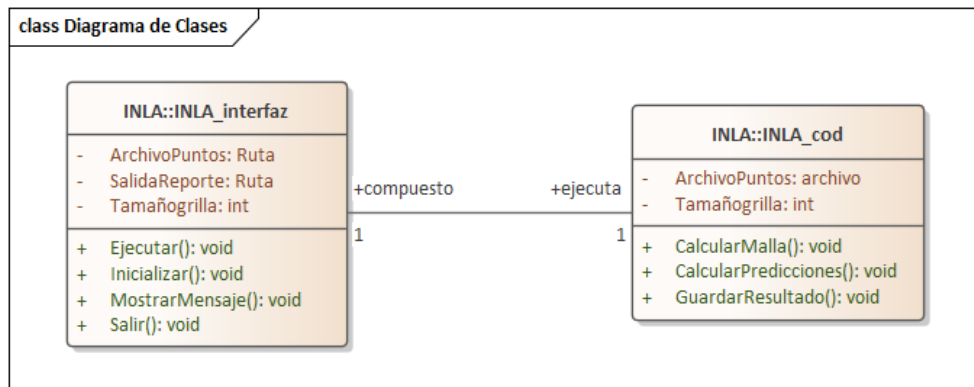


Figura 4: Diagrama de clases

4.2.4. Diagrama de paquetes de clases:

En este diagrama se puede observar la organización y disposición de las clases nombradas anteriormente en forma de paquete según su dependencia, como una clase depende de la otra para la ejecución del plugin se deja un único paquete.

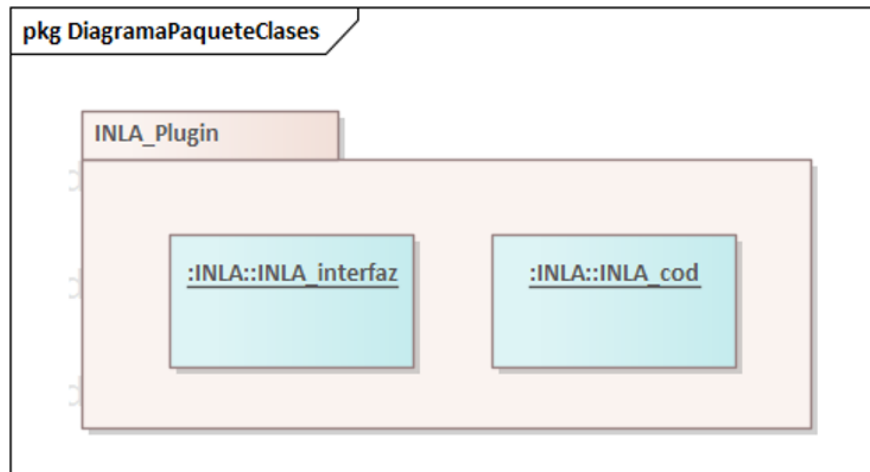


Figura 5: Diagrama de paquete de clases

4.2.5. Diagrama de componentes:

Este modelo permite representar las relaciones entre los componentes individuales que hacen posible el desarrollo del sistema para este caso el plugin funcionará desde el software QGis.

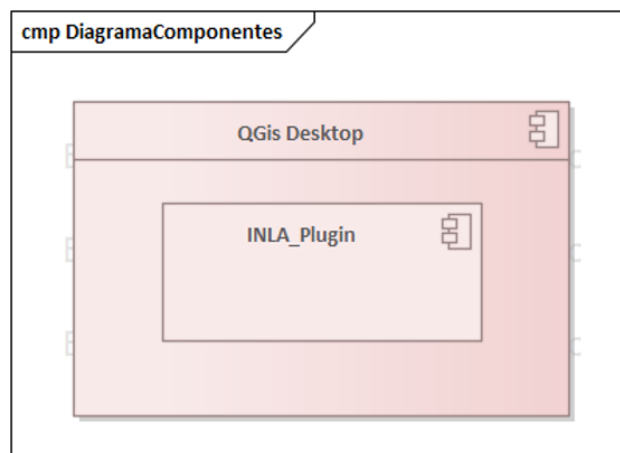


Figura 6: Diagrama de componentes

4.2.6. Diagrama de despliegue:

Este diagrama permite ver la estructura tanto de hardware como de software necesaria para el funcionamiento del complemento, es necesario un computador para ejecutar el instalador de QGIS y desde éste el complemento a desarrollar.

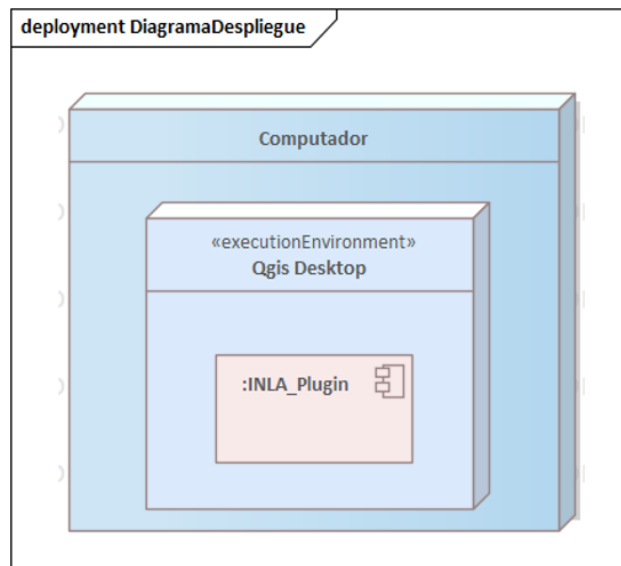


Figura 7: Diagrama de despliegue

4.3. Desarrollo

4.3.1. Herramientas a utilizar

En primera instancia se hace uso del Plugin Builder 3 desde el software QGIS el cual permite estructurar de manera fácil y rápida los archivos que comprenden el paquete de python necesario para el funcionamiento adecuado del complemento a desarrollar, también se hace uso del Plugin Reloader el cual actualiza los cambios que se realicen en el código del complemento desde la interfaz de QGIS y así poder visualizarlos de manera inmediata, en cuanto al diseño de la interfaz gráfica de usuario se hace uso del software Qt Designer el cual crea y diseña cada uno de los ele-

mentos que harán parte de la interfaz del complemento para posteriormente con un editor sencillo como NotePad ++ editar el archivo que contiene la lógica principal y asignarle cada una de las funcionalidades que realizaran dichos elementos.

4.3.2. Diseño de la interfaz gráfica

El usuario deberá instalar el complemento en el software QGIS para poderlo ver como un botón en su barra de herramientas, al pasar el cursor sobre éste le permitirá leer una descripción rápida de lo que provee dicho componente.



Figura 8: Previsualización

Posteriormente el usuario debe cargar sus capas de interés en QGIS para que el complemento pueda leerlas, luego de esto ingresa a la interfaz del complemento y tendrá que completar los campos como se muestra a continuación:

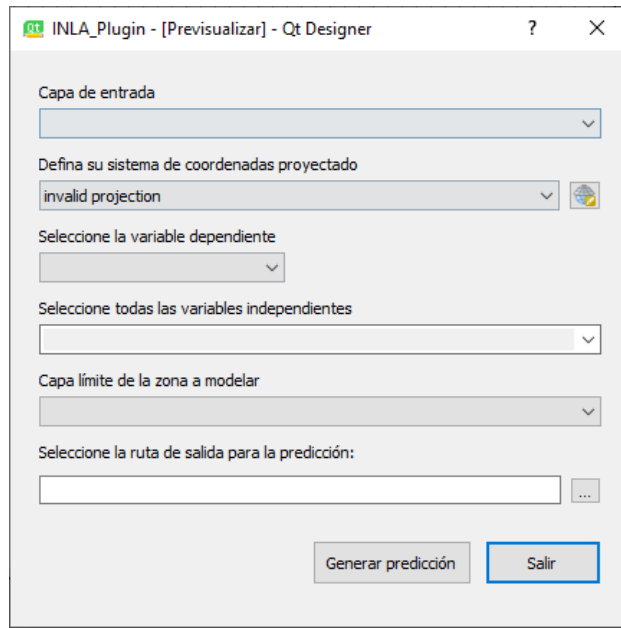


Figura 9: Maqueta del complemento

En primera instancia el usuario debe seleccionar de las capas activas cual es la capa que contiene los puntos muestreados; ésta debe tener geometría tipo punto para sea reconocida por el plugin, si sus datos se encuentran en coordenadas geográficas es necesario que elija un sistema de referencia proyectado para realizar el modelamiento, luego selecciona la variable dependiente o a predecir desde una lista desplegable que tomará los atributos de la capa seleccionada, posterior a esto se despliega una lista con las demás variables que vienen cargadas en la capa para que el usuario seleccione cuales son las variables independientes que participaran como covariables en el modelamiento, después selecciona otra capa tipo polígono que delimita la zona a modelar y por último ubica la ruta donde quiere sea guardado su resultado y puede ejecutar el proceso.

4.3.3. Desarrollo del código lógico

Como se había mencionado anteriormente al ser una metodología de reutilización se procede a descargar el código base del modelamiento en R de la siguiente fuente: <https://github>

.com/JasonRomero11/Modelo-Espacial-Hidrogeologico este código va a ser incorporado a la lógica de cada uno de los componentes de la interfaz desde el lenguaje python lo que le permitirá al usuario realizar el modelamiento de los datos que ingrese.

4.4. Guía de instalación

1. Es recomendable descargar la versión estable 3.18.2-1 del software QGIS para el correcto funcionamiento del plugin, así mismo es necesario tener instalado R en su versión 4.2.1 y de preferencia RStudio en el ordenador de usuario.
2. En la carpeta del plugin se encuentra un archivo llamado “evn_qgis3” el cual debe ser modificado con un editor de texto en la línea 2 y copiar la ruta de la carpeta del software QGIS de su ordenador.
3. Desde la carpeta donde se instaló QGIS se encuentra un archivo por lotes llamado “OS-Geo4W” el cual debe ejecutarse como administrador.
4. Copie la ruta completa del archivo de las variables de entorno que modificó en el paso 2 “evn_qgis3” y péguela en el cuadro de comando que se ejecutó, escriba “python” y verifique que la versión es 3.7.0-
5. Escriba “Exit()” y proceda a instalar los siguientes componentes uno a uno con el comando “pip install rpy2==3.5.2” y “pip install numpy” .
6. En la carpeta del plugin se encuentra el archivo “GDAL-3.0.4-cp37-cp37m-win_amd64.whl” copie la ruta completa del archivo e instálelo desde el cuadro de comandos con “pip install” seguido de la ruta del archivo.

7. Haga la instalación de la misma forma que en el paso anterior con el archivo “Fiona-1.8.13-cp37-cp37m-win_amd64.whl”
8. Con el comando “pip install geopandas” instala el ultimo componente necesario.
9. Descargue la última versión de RGeostats correspondiente a su sistema <http://rgeostats.free.fr/forum/viewtopic.php?f=2&t=48>, desde RStudio : vaya a la pestaña "Paquetes", haga clic en el botón “Instalar”, luego seleccione instalar desde: “archivo de almacenamiento del paquete” y busque el archivo descargado.
10. Instale los paquetes necesarios con el código R que se encuentra en la carpeta del plugin abriendo el archivo desde la carpeta ArchivosComplementarios y presione ejecutar.
11. Desde QGis vaya a complementos/administrar e instalar complementos/ instalar mediante zip y seleccione el archivo inla.zip que se encuentra en la carpeta del plugin, ya podrá verlo en su barra de herramientas, se recomienda instalar plugin reloader para refrescar las funcionalidades del plugin.

4.5. Resultados

En primera medida se tiene como resultado un archivo tipo raster de la zona de estudio con los valores de la variable dependiente escogida por el usuario el cual es cargado automáticamente al proyecto al momento de terminar la ejecución del programa, este archivo cuenta con el sistema de coordenadas proyectado escogido por el usuario.



Figura 10: Resultado espacial del modelamiento

En segunda instancia se crea un archivo pdf con un resumen estadístico de las variables independientes que seleccionó el usuario y unas gráficas que permiten visualmente identificar que variables son mas representativas para el modelo, así mismo se da un valor DIC (criterio de información de desviación) el cual permite comparar entre modelos y escoger así el de mejor ajuste.

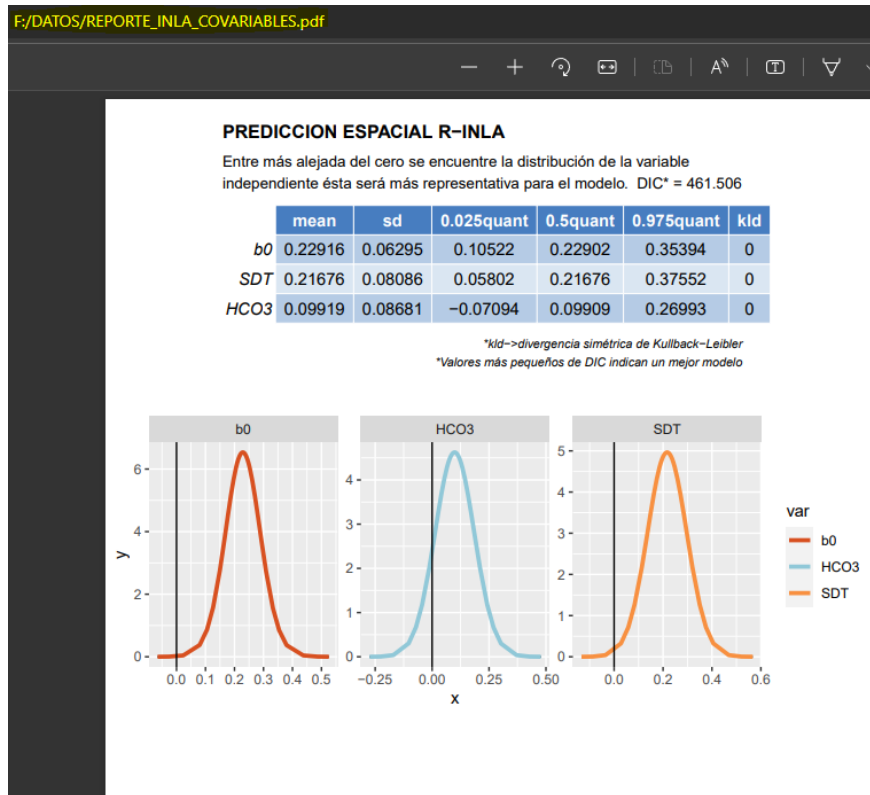


Figura 11: Resultado estadístico del modelamiento

4.6. Pruebas

Dados los tiempos de ejecución del proyecto y en concordancia con el método de muestreo por conveniencia (*No me hagas pensar. Actualización - Anaya Multimedia, 2015*) se seleccionaron 6 profesionales que están altamente relacionados con los SIG y los datos espaciales para que evaluaran la usabilidad, estética de la interfaz y utilidad del complemento, así mismo se registró el tiempo que se demora en ejecutar el modelamiento para el mismo conjunto de datos en cada computador y así evaluar el rendimiento.

4.6.1. Prueba de rendimiento

Con esta prueba se puede evidenciar que en la mayoría de computadores los factores más importantes en el ahorro de tiempo de procesamiento para el modelado espacial con el plugin INLA esta principalmente ligado al número de procesadores lógicos y la velocidad base de la CPU, sin embargo es de resaltar que en general para todos los ordenadores toma tiempos bajos para procesos de interpolación espacial, por lo que se considera puede ser utilizado en cualquiera que cumpla con las especificaciones mínimas de uso en sistemas de información geográfica.

CPU			RAM (GB)	DISCO DURO	TARJETA GRÁFICA	TIEMPO DE EJECUCIÓN (min)
Núcleos	Procesadores lógicos	Velocidad base (GHz)				
4	8	1.6	12	HDD	Interna de 3.9 GB	7
8	8	2	16	SSD	Interna de 8.2 GB	2
4	8	1.8	8	HDD	Interna de 3.9 GB	2.5
4	4	2.7	20	SSD	Interna de 9.7 GB	6
4	8	3.4	8	HDD	Interna de 4.0 GB	1
4	4	2.5	8	SSD	Interna de 4.5 GB	5
2	4	2.6	8	SSD	Interna de 4.9 GB	2.2

Cuadro 1: Resultados de tiempos de ejecución

4.6.2. Usabilidad

Nombre	Profesión	Usabilidad	Diseño de interfaz	Utilidad del complemento
Salomón Ramirez	Ing. Catastral Esp. SIG	3	4	5
Michelle Avila	Ing. Topográfica	4	5	5
Camilo Moreno	Ing. Topográfico	4	5	5
Linda Fajardo	Ing. Ambiental	4.5	5	5
Daniel Hernandez	Ing. Catastral	4.5	4.5	5
Sebastian Sierra	Ing. Civil	3	4.6	5

Cuadro 2: Resultados de experiencia de usuario

Como se puede ver en la guía de instalación se deben realizar varios pasos para instalar paquetes de R, python y configurar variables de entorno; esto redujo la puntuación en temas de usabilidad, una vez instalado el complemento los usuarios a partir de las indicaciones y mensajes del complemento pudieron llevar a cabo exitosamente el modelamiento; sin embargo algunos manifestaron la necesidad de mensajes explicativos sobre el modelo y uso de las variables, finalmente todos concluyeron que el complemento tiene gran utilidad dado el ahorro de tiempos de ejecución para modelamientos espaciales y la facilidad de uso sin tener que adentrarse en temas de programación o estadística espacial.

5. Conclusiones y recomendaciones

5.1. Conclusiones

- Dadas las pruebas de rendimiento realizadas se establece que las características mínimas de un ordenador para el uso del plugin INLA son las mismas básicas que requiere el funcionamiento adecuado de QGIS, es de resaltar que los tiempos de ejecución van a depender

principalmente del procesador y estado general del computador.

- La usabilidad del plugin es el punto que más se debe reforzar en este desarrollo, si bien el rendimiento y la funcionalidad están acorde a los requerimientos funcionales la experiencia de los usuarios debe ser lo más amigable posible.
- Las bondades del lenguaje de programación python permitieron llevar a cabo el desarrollo del complemento al facilitar la integración del lenguaje estadístico R y potenciar sus funcionalidades, permitiendo mostrar gráficos de algunos procesos intermedios para conocimiento del usuario.
- El uso de la metodología INLA fue un acierto ya que al tener costos computacionales bajos pudo ser soportado fácilmente por un sistema de información geográfica como QGis, mejorando la experiencia de los usuarios no expertos con los modelos de predicción espacial.

5.2. Recomendaciones

- Se puede mejorar el tema de usabilidad automatizando los procesos de instalación de las librerías de python y paquetes de R, así como la configuración de las variables de entorno del sistema.
- El plugin puede ser mejorado incluyendo otras metodologías de predicción espacial y hacer comparaciones entre las mismas.
- El complemento podría ser implementado en otros softwares tanto libres como comerciales.

Referencias

- Agudelo, N., Tano, G., y Vargas, C. A. (s.f.). HISTORIA DE LA AUTOMATIZACIÓN. , 5. Descargado de <http://ingenierovizcaino.com/ecci/aut1/cortel/articulos/Historia%20de%20la%20Automatizacion.pdf>
- Análisis y ciencia de datos | identificar patrones y realizar predicciones.* (2022). Descargado 2022-02-04, de <https://www.esri.com/es-es/arcgis/products/arcgis-pro/features/analytics-data-science>
- Arriola, A. E. S. (2021). *La importancia de automatizar procesos en s.i.g.* Descargado 2022-03-11, de <https://es.linkedin.com/pulse/la-importancia-de-automatizar-procesos-en-sig-solis-arriola>
- Banerjee, S., Carlin, B. P., Gelfand, A. E., y Banerjee, S. (s.f.). *Modelado y análisis jerárquico de datos espaciales.* doi: 10.1201/9780203487808
- Blangiardo, M., y Cameletti, M. (2015). *Spatial and spatio-temporal bayesian models with r - INLA.* John Wiley & Sons.
- Bohórquez, I. A., y Ceballos, E. V. (2008). Algunos conceptos de la econometría espacial y el análisis exploratorio de datos espaciales. , 26.
- Briseño, D. M. V. (s.f.). 1.- fundamentos de ingeniería de requerimientos. , 151.
- Casanova, R. (2012). SIGAP: Sistema de información geográfica para el análisis prospectivo. , 10. Descargado de https://www.fig.net/resources/proceedings/fig_proceedings/uruguay/papers/ts04d/TS04D_casanova_sosa_et_al_6451.pdf
- Confiabilidad en el software.* (2016). Descargado 2022-03-12, de <https://ittgweb>

- .wordpress.com/2016/05/29/4-3-confiabilidad-en-el-software/
- Díaz, C. A. P., y Cocunubo, K. S. B. (2021). Nitrógeno, fósforo y potasio en suelos ganaderos en el municipio de Suesca, departamento. , 89.
- Etikan, I., Musa, S. A., Alkassim, R. S., y cols. (2016). Comparison of convenience sampling and purposive sampling. *American journal of theoretical and applied statistics*, 5(1), 1–4.
- Fernández Díaz, M., Espinosa Ramirez, J. G., y García-Jacas, C. R. (2016). Algoritmo paralelo para la interpolación espacial de kriging ordinario. , 10(3), 57–70. Descargado 2022-02-04, de http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S2227-18992016000300005&lng=es&nrm=iso&tlng=es (Publisher: Universidad de las Ciencias Informáticas)
- García, A. C. F., y Castillo, J. N. P. (2019). Técnicas para la predicción espacial de zonas susceptibles a deslizamientos. *Avances: Investigación en Ingeniería*, 16(1), 20–48.
- García León, D., y Beltrán Benavides, A. (s.f.). Un enfoque actual sobre la calidad del software. , 3(3), 40–42. Descargado 2022-03-12, de http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S1024-94351995000300005&lng=es&nrm=iso&tlng=es (Publisher: 2000, Editorial Ciencias Médicas)
- Giannini-Kurina, F., Suarez, F., Paccioretti, P., Macchiavelli, R., y Mónica, B. (s.f.). Desempeño predictivo de r-INLA SPDE para el mapeo digital de suelos. , 10.
- Gómez-Rubio, V. (2020). *Bayesian inference with inla*. CRC Press.
- Havard Rue, N. C., Sara Martino. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. , 71. Descargado de <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2008.00700.x> doi: 10.1111/j.1467-9868.2008.00700.x

- Hilliard, R. (2000). Recommended practice for architectural description of software-intensive systems. , 28.
- Lindgren, F., Rue, H., y Lindström, J. (2011). An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4), 423–498.
- Mejía Rivera, O., Betancur Vargas, T., y Londoño Ciro, L. (2007). APLICACIÓN DE TÉCNICAS GEOESTADÍSTICAS EN LA HIDROGEOLOGÍA DEL BAJO CAUCA ANTIOQUEÑO. , 74(152), 137–149. Descargado 2022-02-04, de http://www.scielo.org.co/scielo.php?script=sci_abstract&pid=S0012-73532007000200012&lng=en&nrm=iso&tlng=es (Publisher: 2006, Revista DYNA)
- Monsalve Graterol, N. C. (2013). *Modelos jerárquicos bayesianos espaciales en epidemiología agrícola* (phdthesis). doi: 10.4995/Thesis/10251/19161
- Moraga, P. (2019). *Geospatial health data: Modeling and visualization with r-INLA and shiny*. Descargado 2022-02-05, de <https://www.paulamoraga.com/book-geospatial-info/book-geospatial-info/>
- Morales, A. (2015). *Automatizar tareas: Model builder o python*. Descargado 2022-02-04, de <https://mappinggis.com/2015/06/model-builder-o-python/>
- Niño Peña, C. A., y Cáceres Castellanos, G. (2014). Estado actual de temáticas para el análisis espacial en la toma de decisiones. , 16.
- No me hagas pensar. Actualización - Anaya Multimedia*. (2015). Descargado 2022-08-03, de <https://anayamultimedia.es/libro/titulos-especiales/no-me-hagas-pensar-actualizacion-steve-krug-9788441537279/>

- Olaya, V. (s.f.). *Historia de los SIG*. Descargado 2022-03-11, de <https://volaya.github.io/libro-sig/chapters/Historia.html>
- Páez, L. O. M. (2011). Descripción general de la infheacreer nincfeiraenBciaas ysoebsreialanaoc-yurrsenucsia de eventos futuros y la relación con el aplicaciones en los protecoreesmoasded-Beayegs eysltaióinnferencia dentro. , 28.
- Reynoso, C. B. (2004). Introducción a la arquitectura de software. *Universidad de Buenos Aires*, 33.
- R-INLA project*. (2022). Descargado 2022-02-05, de <https://www.r-inla.org/>
- Romero Rios, J. M., y Castaño Salazar, D. (2021). MODELADO ESPACIAL HIDROLÓGICO PARA DETERMINAR ÍNDICES DE CALIDAD y VULNERABILIDAD DE LAS AGUAS SUBTERRÁNEAS EN LA ZONA CENTRO DE BOYACÁ. , 147.
- Ruiz, F. (s.f.). Univ. cantabria – fac. de ciencias. , 44.
- Ruiz, M. R. (2020). MAPEO DE ENFERMEDADES CON ENFOQUES BAYESIANOS y EVALUACIÓN DEL RIESGO RELATIVO PARA LA SALUD PÚBLICA EN COLOMBIA. , 99. Descargado de <https://repository.udistrital.edu.co/bitstream/handle/11349/23780/RuizFernandezMarlonRicardo2020.pdf?sequence=1&isAllowed=y>
- Serrano, R. M., y Valcarce, E. V. (2000). *Técnicas econométricas para el tratamiento de datos espaciales: la econometría espacial* (Vol. 44). Edicions Universitat Barcelona.
- Siabato, W. (s.f.). Sobre la evolución de la información geográfica: las bodas de oro de los SIG. , 27(1), 1–9. Descargado 2022-03-11, de <https://revistas.unal.edu.co/index.php/rcg/article/view/69500> doi: 10.15446/rcdg.v27n1.69500
- Statistics, S. (2021). *Inferencia bayesiana sobre modelos de regresión lineal*.

Descargado 2022-03-12, de <https://prod.ibmdocs-production-dal-6099123ce774e592a519d7c33db8265e-0000.us-south.containers.appdomain.cloud/docs/es/spss-statistics/25.0.0?topic=statistics-bayesian-inference-about-linear-regression-models>

Tobler, W. R. (1969). Geographical filters and their inverses. *Geographical Analysis*, 1(3), 234–253.

Valencia Cedeño, A. Y., García Jimenez, D. R., Ascencio Arias, J., y cols. (2016). Qbiogridmonitor: Plugin en qgis para la determinación de áreas de muestreos biológicos.

¿qué es la automatización? (2022). Descargado de <https://www.redhat.com/es/topics/automation/whats-it-automation>

¿qué es la usabilidad? | guía digital. (2022). Descargado de <https://www.guiadigital.gob.cl/articulo/que-es-la-usabilidad.html>

¿qué es un plugin y para qué se usa? (2020). Descargado 2022-03-11, de <https://www.ionos.es/digitalguide/servidores/know-how/que-es-un-plugin/>