

García Ginés, J., Donaire Apala, S.C. y Gómez Delgado, M. (2010): Validación de los resultados de los análisis llevados a cabo en un SIG ráster: cálculo de camino óptimo. En: Ojeda, J., Pita, M.F. y Vallejo, I. (Eds.), *Tecnologías de la Información Geográfica: La Información Geográfica al servicio de los ciudadanos*. Secretariado de Publicaciones de la Universidad de Sevilla. Sevilla. Pp. 482-497. ISBN: 978-84-472-1294-1

VALIDACIÓN DE LOS RESULTADOS DE LOS ANÁLISIS LLEVADOS A CABO EN UN SIG RÁSTER: CÁLCULO DE CAMINO ÓPTIMO

José García Ginés¹, Susan Cinthia Donaire Apala² y Montserrat Gómez Delgado³

Departamento de Geografía. Universidad de Alcalá. C/ Colegios, 2. 28801 Alcalá de Henares, Madrid.

(1) multiatleta@hotmail.com; (2) susan_cinthia@hotmail.com; (3) Montserrat.gomez@uah.es

RESUMEN

La aplicación de procesos de validación que aporten cierta robustez y confianza a los resultados obtenidos de los análisis y modelos ejecutados en un entorno SIG (Sistemas de Información Geográfica) sigue siendo una asignatura pendiente.

En el presente trabajo se pretende comprobar si la metodología de validación propuesta por Gómez Delgado y Barredo (2005) aplicada a un proceso de Evaluación MultiCriterio(EMC) llevado a cabo en un entorno SIG, puede ser extendida a otros de los análisis habitualmente llevados a cabo en este mismo entorno, y cuyos resultados pueden tener cierta trascendencia también en la toma de decisiones, como es la determinación de la trayectoria y construcción de nuevas infraestructuras en el territorio. Así, se ha realizado un ensayo en dos zonas de la Comunidad de Madrid, aplicando a los resultados del análisis de cálculo óptimo un proceso de validación basado en la combinación de un análisis de sensibilidad y un análisis de incertidumbre. Con esta metodología se pretende comprobar la robustez de los resultados y el impacto que pudiese tener la existencia de errores en los datos de partida.

Palabras Clave: Análisis de incertidumbre, análisis de sensibilidad, SIG, cálculo de camino óptimo, Madrid.

ABSTRACT

The application of validation processes that can provide robustness and reliability to results obtained from GIS analysis and models is still needed.

The main purpose of the work presented here was to test whether the validation methodology proposed by Gómez Delgado and Barredo (2005), applied to a Multicriteria Evaluation (MCE) process carried out in a GIS environment, could be extended to other GIS analysis whose results could also have an impact in decision making, for instance in regards to the determination of the trajectory and construction of new infrastructures in the territory. Thus, a test was undertaken in two areas of the Comunidad de Madrid (the Región of Madrid). A validation process based on the combination of sensitivity and uncertainty analyses was applied to the results of an optimal computation analysis. With this methodology we aim at verifying the robustness of results and the eventual impact that errors in the ingested data could have.

Key Words: Uncertainty analysis, sensitivity analysis, GIS, least-cost pathway, Madrid.

1. INTRODUCCIÓN

Los nuevos modelos territoriales surgidos del crecimiento urbano experimentado en los últimos años implican un incremento sustancial del número de desplazamientos, no sólo de pasajeros, sino también de mercancías (Gutiérrez Puebla y García Palomares, 2005). Según el estudio realizado por Seguí Pons y Martínez Reynés (2003), el transporte ferroviario es el tercer medio de transporte más utilizado en España, destacando la instauración y extensión de las redes de alta velocidad entre las grandes regiones metropolitanas, así como el incremento constante en el servicio de mercancías. También destacan el crecimiento sorprendente de las líneas de alta velocidad en Europa en los últimos veinte años, siendo España uno de los ocho países que más kilómetros de vía ha desarrollado. No obstante, hay que señalar que seguimos siendo uno de los países que más volumen de mercancías trasladamos por carretera. Concretamente en la Comunidad Autónoma de Madrid (área en la que se lleva a cabo nuestro estudio), el 97,3% de las entradas de mercancías y el 99,2% de las salidas, se realizaron por carretera durante el año 2006 (Santiago, 2008:160). En este contexto, la creación de nuevas infraestructuras de transporte se ha convertido en una necesidad habitual de nuestra sociedad. Debido a las implicaciones ambientales y territoriales que conlleva la construcción de éstas, resulta imprescindible realizar una tarea exhaustiva de planificación que incluya los análisis y cálculos necesarios para determinar la dirección y el territorio que deben o pueden ocupar.

En la última década, las Tecnologías de la Información Geográfica (TIG) en general y los SIG en particular, se han consolidado como herramientas indispensables en las tareas de planificación territorial, y concretamente en la toma de decisiones relativas a la planificación del transporte en sus diferentes modalidades (Backhoff Pohls, 2005). Dada la importancia que han adquirido sus aplicaciones en esta área, en el ámbito anglosajón se le ha dado una nomenclatura específica conocida como SIG-T (*GIS-T*, *GIS for Transport*), integrando en este grupo tanto procesos de manipulación y análisis de datos, como de modelización (que no siempre vienen incluidos en los SIG convencionales).

En este contexto, uno de los análisis frecuentemente utilizados es el cálculo de rutas óptimas. Sin embargo no se ha prestado la debida atención a la validación de los resultados que obtenemos con estos procedimientos. La realidad es que la información geográfica utilizada en un SIG y los análisis realizados con ella, se ven afectados por la incertidumbre, entendida como cualquier error conocido o no, ambigüedad o variación en una base de datos y en los modelos de análisis. Esta falta de certidumbre o de seguridad puede poner en duda la validez de los resultados de esos análisis y operaciones (Gómez Delgado y Bosque Sendra, 2004). El problema es que no se cuenta todavía con procedimientos suficientemente flexibles y fiables para determinar el grado de error final de un proceso SIG relativamente complejo, donde se utilizan numerosos datos y diferentes procedimientos analíticos. Mas aún, la dificultad principal es que, para poder hacer un análisis de propagación de los errores iniciales, se necesita conocerlos bien y de forma cuantitativa, y esto tampoco es habitual hasta el día de hoy.

El objetivo de este trabajo es aplicar una serie de métodos de validación, de sencilla aplicación en un entorno SIG, para dar robustez a los resultados obtenidos a partir de este tipo de análisis. Para ello partimos de la metodología propuesta por Gómez Delgado y Barredo (2005), aplicada a un problema de planificación resuelto mediante técnicas EMC y en un entorno SIG, adaptándola a las características propias del cálculo de rutas de mínimo coste.

En la primera parte del trabajo se realizará una revisión sobre el estado de la cuestión en cuanto a la aplicación del cálculo de camino óptimo y la propuesta de procesos de validación. Tras la presentación del área de estudio y una breve descripción del procedimiento utilizado para obtener las dos trayectorias, se desarrollará el procedimiento de validación y se expondrán los principales resultados obtenidos.

2. MARCO TEÓRICO

Desde el punto de vista teórico, podemos decir que la determinación de trayectorias de vías de comunicación a partir del cálculo de camino óptimo, estaría englobada en el área de aplicación del SIG-T relacionada con la gestión y planificación de carreteras, autopistas, vías férreas y logística de transporte.

Como afirman Yildirim y otros (2006), este tipo de operaciones requiere medidas eficaces de recopilación, tratamiento, almacenamiento y análisis de datos espaciales, tales como la topografía, vegetación, geología, tipo de suelo, el uso de la tierra, deslizamientos de tierra, etc., por lo que el SIG es considerado la mejor plataforma de acción. El manejo de esta tecnología ha cambiado la manera de calcular la trayectoria a seguir de las infraestructuras lineales (carreteras, vías férreas, etc.), enriqueciendo de manera considerable los métodos clásicos que se basaban en una menor cantidad de datos (topográficos principalmente).

Sin embargo, no encontramos en la literatura muchos ejemplos que utilicen este tipo de aplicaciones, aunque hemos de reconocer que son bastante variadas como veremos a continuación. Gómez Delgado y otros (1995), desarrollan una metodología basada en la combinación de costes de construcción y costes ambientales para diseñar una nueva carretera en el noreste de la Comunidad de Madrid que minimice los costes de desplazamiento. El estudio pretende mostrar las posibilidades de generar varias alternativas que respondan a la priorización de distintos criterios (ambientales o meramente económicos) para que la toma de decisiones se lleve a cabo de manera más informada. López Romero (2005), utiliza este procedimiento en el ámbito de la investigación arqueológica, con el objetivo de estimar cuáles habrían sido las rutas de comunicación entre distintos yacimientos arqueológicos de la misma época en la Segeda (Zaragoza). Para ello utiliza una fricción basada en la topografía, la distancia a cursos de agua y la vegetación. Feldman y otros (1995), determinan el trazado de 700km de oleoducto en la región del mar del Caspio, teniendo en cuenta factores topográficos, la geología, los usos del suelo, la red hidrográfica y las vías de comunicación existentes. El estudio llevado a cabo por Delavar y Naghibi (2003) es un ejemplo muy parecido al anterior, pero en esta ocasión en el sudeste de Irán y con una longitud de sólo 34km. Una aplicación completamente diferente la encontramos en el trabajo de Ganskopp y otros (2000). Estos autores comprueban que las rutas habitualmente seguidas por el ganado desde las zonas de estabulación a las de pasto son las de mínimo esfuerzo. Para ello utilizan, básicamente, la pendiente y la orientación del terreno. Finalmente, Dane y Tecim, (2007), proponen integrar SIG y EMC para la construcción de una nueva vía de ferrocarril que cumpla unos criterios mínimos de sostenibilidad. Para ello restringen el área de acción, combinando con operaciones booleanas una serie de criterios ambientales y socioeconómicos, y una vez obtenida el área de estudio susceptible de ser utilizada para este propósito, obtienen las áreas más adecuadas mediante un proceso de EMC (concretamente el método de las Jerarquías Analíticas de Saaty) que combina esos mismos factores. Hemos de señalar, sin embargo, que en este caso no se utiliza el análisis de cálculo de camino de mínimo coste apropiadamente dicho.

Como hemos mencionado ya en la introducción, la información geográfica no está exenta de error. Existen diversos estudios relativos al origen, cuantificación y modelización de la propagación de estos errores (Haining et al., 1998; Heuvelink, 1998; Hunter, 1999). Sin embargo el proceso de caracterización del error es bastante complejo, y las posibilidades y herramientas para poder realizar procesos de verificación completos son bastante limitadas (Gómez Delgado y Bosque Sendra, 2004).

En este contexto, es más habitual que, en lugar de un proceso de verificación, se recurra a un proceso de validación cuyo objetivo es determinar si un modelo y los resultados obtenidos mediante la aplicación del mismo son aceptables para el uso que se les pretende dar por ajustarse a los requisitos de actuación (Rykiel, 1996). Sin embargo, validar los resultados de los modelos empleados para asistir la toma de decisiones en labores de planificación no es una tarea comúnmente extendida, limitándose, en muchos casos, a un ejercicio de comprobación de los resultados sin otorgarles ningún grado de garantía. Dada la trascendencia de las decisiones territoriales que se puedan tomar basadas en los resultados de la aplicación de un determinado modelo, parece necesario que tanto el modelo empleado como los resultados que proporciona, gocen de una cierta robustez (Gómez Delgado y Bosque Sendra, 2004).

A pesar de esto, es habitual que los problemas de error e incertidumbre que pueden presentar los datos utilizados en un SIG o los efectos que puedan tener sobre los resultados obtenidos de los modelos implementados en ellos sean ignorados (Gómez Delgado y Barredo, 2005). Además, se debe tener en cuenta que los errores se propagan en cada una de las operaciones que se realizan con la información empleada, lo que incrementa aún más la incertidumbre de la validez de los resultados obtenidos y de las conclusiones derivadas de los mismos (Haining et al., 1998).

Según Crosetto y Tarantola (2001), la incertidumbre debe ser evaluada para dotar de confianza las decisiones que se deriven de los resultados obtenidos de la implementación de los modelos. El análisis de incertidumbre y el análisis de sensibilidad cubrirían aspectos complementarios. El análisis de incertidumbre se enfoca a evaluar cómo se propaga a través del modelo la incertidumbre existente en los datos de partida o en los parámetros del modelo y como afecta ésta a los resultados. El análisis de incertidumbre se puede considerar como un test de calidad del modelo y de los datos de entrada, considerando simultáneamente todas las fuentes de incertidumbre. Por otro lado, el análisis de sensibilidad determina como contribuye individualmente cada fuente de incertidumbre en la incertidumbre total del resultado (Crosetto y Tarantola 2001).

No existen muchos estudios espaciales que aborden ambos análisis de manera simultánea. Crosetto y Tarantola (2001) proponen un procedimiento para caracterizar la incertidumbre de datos raster para la detección de área quemada a partir de datos NOAA AVHRR, mostrando la sinergia existente entre el análisis de incertidumbre y el análisis de sensibilidad. Gómez Delgado (2003) proponen una aplicación del análisis de incertidumbre y un análisis

de sensibilidad como medio para validar la localización de una instalación no deseable obtenida mediante la utilización conjunta de los análisis propios de un SIG y EMC.

En este último estudio citado, ya se realizaba una reflexión sobre la problemática del tratamiento del error y de las escasas ejemplos existentes en la literatura de metodologías que fueran suficientemente generalistas y flexibles, como para poder ser aplicadas en distintos contextos y a los diferentes análisis que pueden realizarse en un SIG. Hemos de decir que, a día de hoy, la situación no ha cambiado en exceso. De hecho hemos de advertir que en ninguno de los estudios mencionados en los que se aplica el cálculo de camino óptimo se hace alusión alguna o se trata el problema del error de los datos de partida o de su propagación, o a la estabilidad o robustez de los resultados.

3. OBTENCIÓN DE UNA NUEVA TRAYECTORIA DE CARRETERA MEDIANTE EL CÁLCULO DE CAMINO ÓPTIMO. DOS SIMULACIONES EN LA COMUNIDAD DE MADRID

Como hemos mencionado, la validación se realizará sobre dos ensayos en la fase de anteproyecto de trazado de dos tipos de infraestructuras: una carretera y un tramo de ferrocarril. Ambos ejemplos están justificados por las características de las zonas de estudio de la Comunidad de Madrid elegidas, con un importante potencial de desarrollo que puede crear necesidades de ampliación de infraestructuras en un futuro cercano, pero que todavía cuentan con una componente rural muy importante.

En el primer caso hemos tratado de reproducir el mismo análisis realizado en un trabajo anterior ya comentado en la sección anterior (Gómez Delgado et al., 1995), en el área de Torrelaguna, al Norte del municipio de Madrid. En el segundo caso se plantea la construcción de una nueva vía de ferrocarril que uniera un posible futuro aeropuerto en el municipio de Campo Real y la estación de ferrocarril de Atocha (en el centro de Madrid capital) (Figura 1). Este segundo caso responde a una necesidad si cabe más fundamentada, pues ya se han realizado hasta 36 estudios para la ubicación en esta zona del futuro aeropuerto (es posible encontrar más información en: <http://www.madridiario.es/2007/Noviembre/transporte/45281/plan-infraestructuras-aeropuertarias-madrid.html>).



Figura 1. Localización de las dos áreas de estudio en la Comunidad de Madrid

La metodología propuesta para la búsqueda de una ruta óptima será la habitualmente utilizada en este tipo de análisis. Se trata de crear estratos temáticos de cada uno de los criterios de adecuación, los criterios se convierten a valores de fricción unitaria (representando la dificultad de paso de la ruta por cada punto), para luego ser combinadas y de esta manera obtener una superficie de costes acumulados para llegar a uno de los extremos de la ruta. Finalmente se determinará por dónde ha de pasar la ruta de manera que se minimice el costo acumulado.

La información de partida para configurar las distintas variables de tipo ambiental y económico que serán utilizadas como determinantes de la fricción que presenta el territorio provienen, en su mayoría, de distintas conserjerías de la Comunidad de Madrid. Por otro lado, la escala original de la información utilizada es mayoritariamente 1:200 000. Dadas las características del análisis a ser realizado y la flexibilidad propia del entorno raster, toda la información ha sido transformada a este formato utilizando un tamaño de píxel de 50m para la zona de Torrelaguna y 200m para el caso de la base de datos del área de estudio de Campo Real. El SIG utilizado para realizar el presente análisis ha sido IDRISI Andes.

Como hemos mencionado, reproducimos para el trazado de la nueva carretera (**caso 1**) entre Colmenar Viejo y Venturada, el trabajo realizado por Gómez y otros (1995). En este caso se introduce alguna información adicional para evitar que la nueva trayectoria atravesase espacios naturales protegidos o instalaciones especiales como la cárcel de Soto del Real. De manera resumida el objetivo ha sido reducir los costes de construcción, evitando aquellas zonas donde la dureza del material litológico desaconseje la realización de la obra; favorecer que el nuevo trazado aproveche algunos tramos de carreteras ya existentes (pero de orden jerárquico menor); evitar las zonas de excesiva pendiente o reducir, en la medida de lo posible, la construcción de puentes elevados para salvar la circulación de cursos fluviales. Desde el punto de vista ambiental se pretende preservar aquellas zonas de mayor valor ecológico (formaciones boscosas autóctonas, principalmente).

Para el caso de la nueva vía de ferrocarril (**caso 2**), se ha intentado alcanzar prácticamente los mismos objetivos con variables similares. Dado que el procedimiento de cálculo es idéntico para ambas infraestructuras, en la tabla 1 especificamos solamente el primer caso como ejemplo, con las variables utilizadas como fricción y el coste unitario otorgado a cada una de ellas.

A partir de estas premisas, y siguiendo el procedimiento habitual en estos análisis, se calculó el mapa final de fricción, con éste y el punto de origen de la ruta se obtuvo el mapa de costes acumulados y, a partir de ahí, la trayectoria final. En la figura 2 aparece como ejemplo las fases de ejecución del caso 2 y en la figura 3 la ruta óptima obtenida.

Tabla 1. Valor de coste unitario asignado a las distintas categorías de cada variable utilizada en el cálculo de la nueva carretera entre Colmenar Viejo y Venturada.

VARIABLES	CATEGORÍAS	COSTES
LITOLOGÍA		
<i>Materiales compactos</i>	Granitos, neises, cuarcitas, calizas	5
<i>Disgregables</i>	Pizarras	3
<i>Poco coherentes</i>	Arcillas, arenas, margas y yesos	1
CARRETERAS		
	Autovías, nacionales, principales	100
	Urbanas	100
	No existe carretera	5
	Caminos	3
	Secundarias y locales	1
PENDIENTE		
	Mayor 10%	100
	5%-10%	5
	3%-5%	3
	0%-3%	1
HIDROGRAFÍA		
	Embalse	100
	Canal	5
	Ríos y arroyos	3

VARIABLES	CATEGORÍAS	COSTES
	Sin red fluvial	1
USOS DEL SUELO		
<i>Vegetación</i>	Vegetación de ribera	4
	Encinares y sabinares	4
	Pinares	2
	Cultivos agrícolas de regadío	2
	Cultivos agrícolas de secano	1
	Mosaico de cultivos de secano	1
	Matorrales y pastizales	1
<i>Urbano</i>		100
PARQUES NATURALES		
	Red de espacios protegidos	100
	Resto de superficie	1
INFRAESTRUCTURAS		
	Cárcel	100
	Líneas eléctricas, gaseoducto	3
	Resto de superficie	1

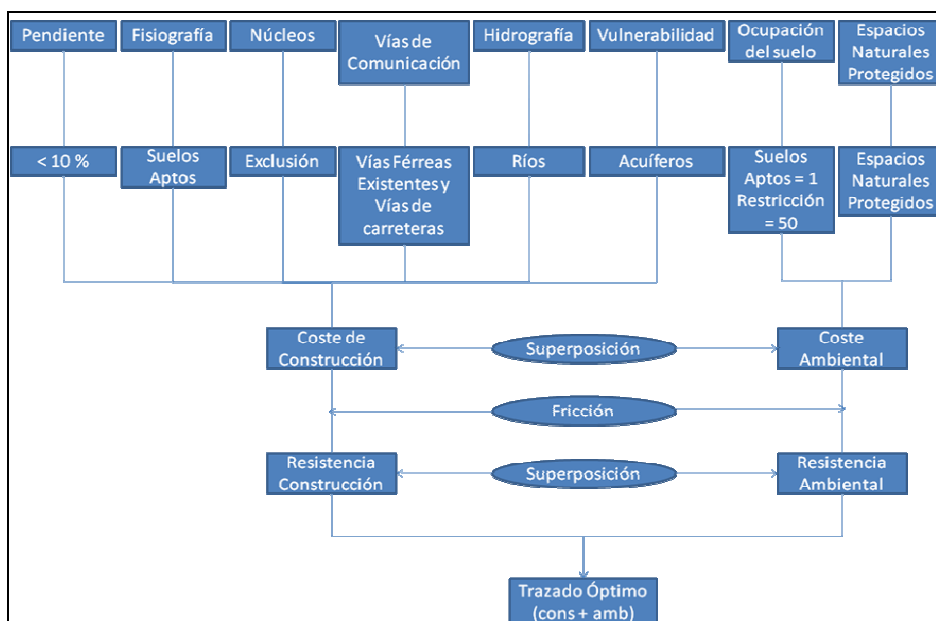


Figura 2. Esquema metodológico para el cálculo de una nueva vía de ferrocarril entre Campo Real y Madrid

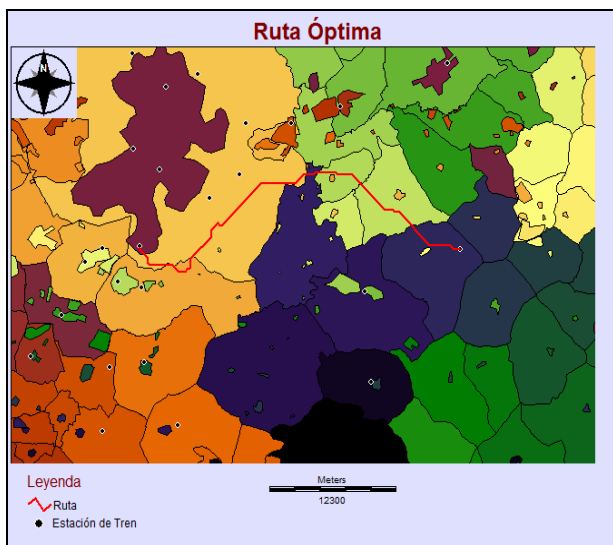


Figura 3. Trayectoria resultante para el caso de la nueva vía de ferrocarril entre Campo Real y Madrid

4. PROCESO DE VALIDACIÓN: METODOLOGÍA Y RESULTADOS

Como ya mencionamos, el objetivo del trabajo es intentar comprobar si es posible aplicar el proceso de validación propuesto por Gómez Delgado (2003) a otro tipo de análisis habitualmente utilizado en un entorno SIG y de gran interés en tareas de planificación, como lo es la determinación de las trayectorias más adecuadas para la construcción de nuevas infraestructuras de transporte. Para ello se han realizado distintas pruebas, adaptando el proceso metodológico a las condiciones específicas del tipo de análisis aquí tratado.

Dicho proceso de validación pasaría por la ejecución de dos tipos de análisis: uno de sensibilidad y otro de incertidumbre, que permitirían dar mayor fuerza o robustez a los resultados obtenidos, y, por otro lado, controlar el riesgo que implicaría tomar cualquier decisión basada en los mismos (figura 4). Pensamos que se trata de una metodología atractiva por su fácil implementación, pues se basa en la utilización de las herramientas básicas de las que dispone cualquier SIG comercial, sin necesidad de programar ni desarrollar módulos específicos para este fin. No obstante, y como cualquier otro procedimiento que requiere ejecutar de manera reiterada un mismo proceso con distintos escenarios de partida, implica la producción de un gran volumen de datos que hay que controlar y gestionar correctamente, lo que requiere una determinada capacidad de procesamiento de los datos para que la ejecución de los procesos no se hagan demasiado largos, y así no desanimar al usuario.

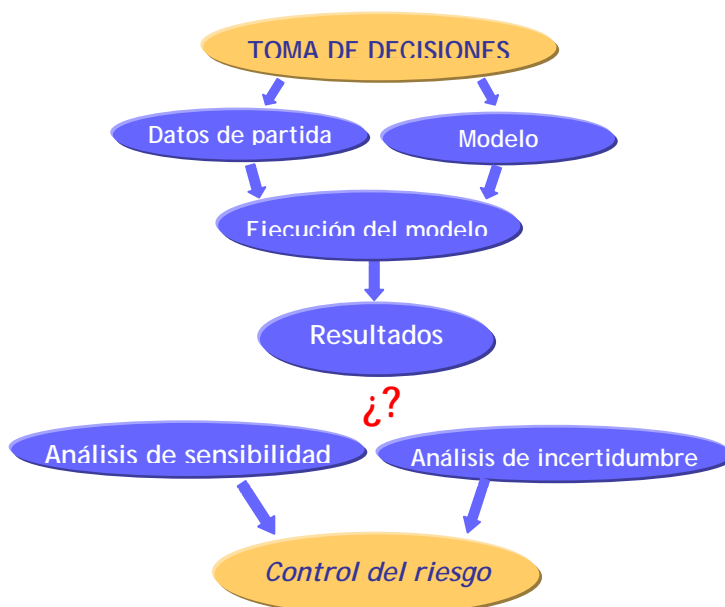


Figura 4. Combinación de análisis de sensibilidad y análisis de incertidumbre como proceso de validación, y por tanto, de control del riesgo, en un proceso de toma de decisiones.

4.1 Análisis de sensibilidad

El objetivo era examinar si las variaciones realizadas en un rango de interés sobre uno o varios parámetros de entrada causaban variaciones importantes en los resultados. En este caso, y dada la forma de aplicación de los algoritmos necesarios para el cálculo de la ruta óptima, no fue posible utilizar análisis de sensibilidad como el de Sobol' o FAST, habitualmente utilizados en otros campos científicos y que habían dado buenos resultados en el ámbito de aplicación de la EMC en un entorno SIG (Gómez Delgado y Tarantola, 2006). Así, y siguiendo el principio esencial de este tipo de análisis, se realiza una adaptación del mismo al procedimiento del cálculo de camino óptimo, desarrollando dos propuestas diferentes, cada una aplicada a un área de estudio.

En un primer ensayo se decidió no considerar igual de importantes todas las variables que participaban en el cálculo de la superficie de fricción del **caso 1**. Así, se otorgó un peso a cada una de ellas mediante el método de la matriz de comparaciones de Saaty, excluyendo los espacios naturales protegidos pues, como podemos comprobar en la tabla 1, actúan prácticamente de restricción. En este caso se optó por otorgar una mayor importancia a los aspectos ambientales, intentando que la trayectoria evitase atravesar las zonas de especial interés ecológico. La ponderación de las distintas variables quedó finalmente como sigue:

Vegetación: 0,43	- Litología: 0,09
Pendiente: 0,2	- Carreteras: 0,05
Hidrografía: 0,2	- Instalaciones: 0,026

Generada la nueva superficie de fricción, se procedió a calcular de nuevo el mapa de costes acumulados y la trayectoria final. El análisis de las variaciones que esta modificación podía provocar se realizó mediante dos métodos. En primer lugar analizamos las coincidencias entre los dos mapas de costes acumulados a partir de un análisis de regresión por píxel. Este modelo matemático establece la relación entre una variable dependiente Y, la variable independiente X y un término aleatorio. Por su simplicidad analítica es la relación lineal la forma funcional que más se utiliza. El coeficiente de determinación R^2 de las rectas informará de la calidad del ajuste, siendo bueno si R^2 es cercano o igual a 1.

Un coeficiente R^2 de 0,96, claramente pone de manifiesto que las diferencias introducidas por la asignación de pesos a cada una de las variables influyen muy poco o casi nada en la obtención del mapa de costes acumulados final. Sin embargo, comparando las dos trayectorias resultantes (figura 5), observamos que, pese a que el resultado puede considerarse bastante robusto en la mayor parte del recorrido, se registran diferencias de cierta importancia en dos tramos (al norte y al sur).

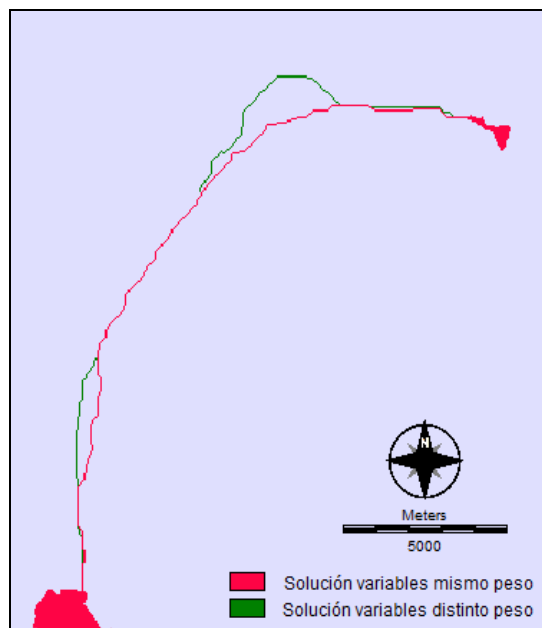


Figura 5. Comparación de las rutas obtenidas para el caso 1 (carretera entre Colmenar Viejo y Venturada), con y sin asignación de pesos a las variables utilizadas para calcular la superficie de fricción

En un segundo ensayo aplicado en el **área de estudio 2**, se procedió a modificar sensiblemente el valor de coste unitario otorgado a las categorías más importantes de las variables que conformarían la superficie de fricción final. Concretamente se procedió a incrementar su valor en, aproximadamente, un 35%. Una vez realizadas las modificaciones, se procedió a generar hasta siete mapas de fricción diferentes, introduciendo en cada ocasión algún factor modificado. A partir de ahí se obtuvieron siete mapas de coste diferentes y siete nuevas trayectorias.

En primer lugar, del análisis de regresión por píxel de los siete mapas de costes acumulados modificados respecto al original, se deriva que los resultados son bastante estables, pues los coeficientes R^2 oscilan entre 0,93 y 0,99. En este caso, y siguiendo la propuesta de Gómez Delgado y Bosque Sendra (2004), se analizaron las diferencias encontradas entre todas las trayectorias obtenidas y la original, a partir de la realización de una cartografía de los píxeles más veces seleccionados. Como podemos comprobar en la figura 6, las coincidencias son importantes en la mayor parte de la trayectoria, no habiendo ninguna parte de la misma que no haya sido seleccionada menos de 5 veces. No obstante, en las proximidades de Campo Real se produce una acumulación de las menores coincidencias, lo que debería ser objeto de un análisis más pormenorizado de las variables que están influyendo en esta variación.

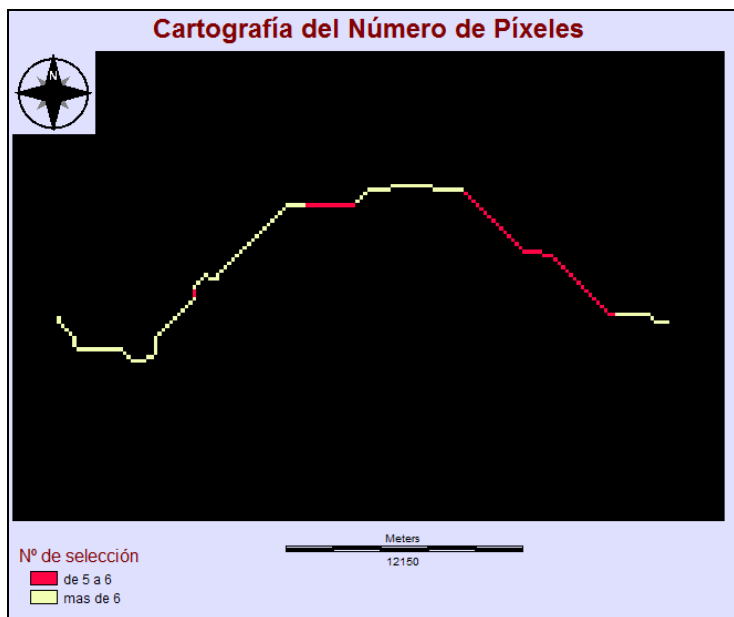


Figura 6. Cartografía de los píxeles más veces seleccionados en las siete ejecuciones realizadas para el caso 2 (línea de ferrocarril entre Madrid y Colmenar Viejo).

4.2 Análisis de incertidumbre

Para la realización del análisis de incertidumbre se siguió la misma metodología propuesta por Gómez Delgado y Bosque Sendra (2004). Los buenos resultados obtenidos con la misma aconsejan su utilización cuando llevamos a cabo análisis en un entorno SIG raster, como es nuestro caso. El objetivo es demostrar que su aplicación también resulta sencilla y válida para el cálculo de camino óptimo.

De manera resumida podemos decir que este análisis se basa en la simulación de la existencia de errores en los datos geográficos de partida (tanto en la componente espacial como en la componente temática), y en la comprobación (mediante una simulación de Monte Carlo) del alcance que estos errores y la propagación de los mismos a través de las operaciones realizadas en el SIG tienen en los resultados del análisis o procedimiento realizado. Hemos de advertir que esta práctica de simular errores o distorsiones en los datos de entrada ha sido también utilizada en varios trabajos, fundamentalmente centrados en temas de cambios de ocupación de uso del suelo (Yeh y Li, 2006; Burnicki et al., 2007; Rae, et al., 2007).

4.2.1 Distorsión de la componente espacial y la componente temática en las dos áreas de estudio.

El procedimiento de distorsión de la **componente espacial** de los datos utilizados en el estudio se basa en la aplicación de un procedimiento análogo a la corrección geométrica, pero con el objetivo de distorsionar en lugar de georeferenciar correctamente los datos espaciales. Una vez seleccionados una serie de puntos de control en el mapa, las coordenadas X e Y serán aumentadas o disminuidas de manera aleatoria y en una cantidad también aleatoria. En nuestro caso se realizaron distorsiones en tres intervalos de 0-250m, 0-500m y 0-700m.

El procedimiento de distorsión de la **componente temática** se efectúa aleatoriamente sobre una cierta cantidad de píxeles de toda la superficie de estudio en cada una de las variables consideradas. Para ampliar las alternativas en el proceso de análisis, y siguiendo la metodología mencionada, se establecieron dos escenarios. En el primero se seleccionaron aleatoriamente hasta un 2% del número de píxeles que conforman el área de estudio y en el segundo hasta un 4%. A la hora de realizar la distorsión, el proceso variará según se trate de una variable cualitativa o cuantitativa. En el primer caso se realizará un cambio aleatorio del identificador de cada categoría en los píxeles seleccionados. En el caso de las variables cuantitativas, se cambiará el valor original por otro aleatorio dentro del rango del 40% del valor máximo que alcanza la variable original. En la figura 7 y 8 encontramos los dos procesos de

distorsión esquematizados y acompañados de un ejemplo (el esquema detallado de ejecución de ambos procesos en el SIG Idrisi puede encontrarse en Gómez Delgado y Bosque Sendra (2004), apéndices 1 y 2).

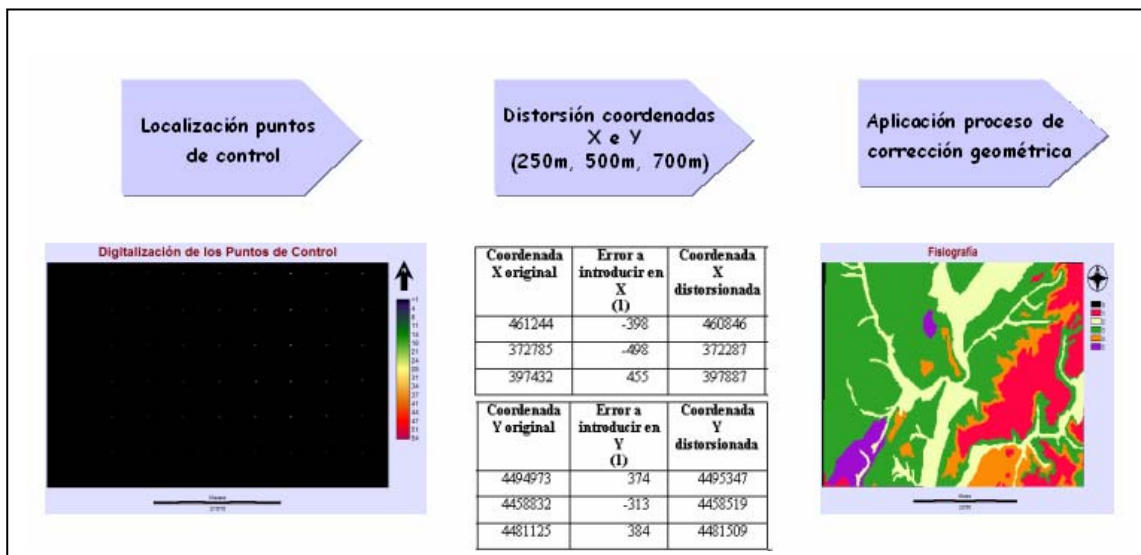


Figura 7. Esquema del proceso de distorsión de la componente espacial

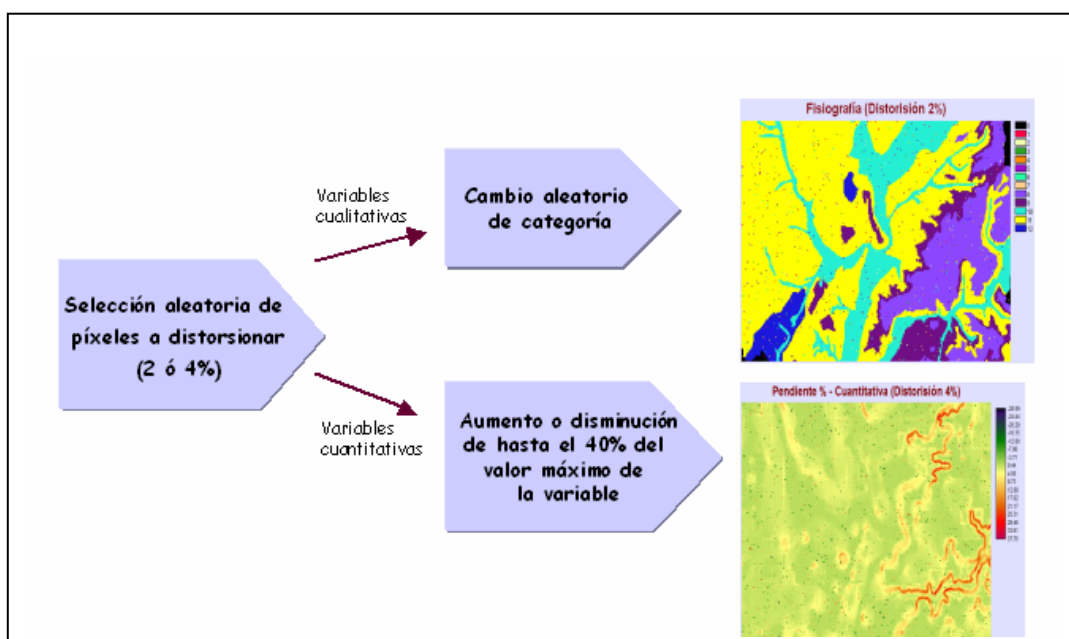


Figura 8. Esquema del proceso de distorsión de la componente temática

Llegados a este punto, contamos con 4 nuevas versiones de cada una de las variables originales implicadas en el estudio. El siguiente paso consistirá en ejecutar el cálculo de camino óptimo un número significativo de veces, utilizando un proceso análogo al análisis de Monte Carlo. Para ello se introdujo una variable distorsionada en cada simulación, manteniendo para el resto la información original. En un segundo momento se realizaron una serie de simulaciones introduciendo todas las variables con algún tipo de distorsión. Si bien hemos de reconocer que considerar que todas las variables implicadas tienen algún tipo de error no resulta muy realista, pensamos que este tipo de simulación podría ofrecernos una seguridad definitiva sobre la robustez de los resultados.

Considerando que en el caso 1 disponíamos de 7 variables, el número de simulaciones con una de las variables distorsionadas ascendió a 24, a lo que se sumaron 8 simulaciones más con todas las variables con algún tipo de distorsión (espacial o temática). En el caso 2 se trabajaba con 4 variables, por lo que el número de simulaciones del primer tipo ascendía a 16, a las que se añadieron 9 iteraciones con todas las variables distorsionadas. El resultado de la ejecución de este proceso supone generar 32 y 25 mapas de fricción, 32 y 25 mapas de costes acumulados y 32 y 25 mapas con la ruta óptima final.

4.2.2 Análisis de los resultados.

El proceso de validación se realizó sobre dos planos de obtención de resultados diferentes, lo que dota a este procedimiento de una solidez aún mayor. Por un lado se analizaron los mapas de costes acumulados obtenidos tras las simulaciones y, por otro, los resultados de la ruta óptima. De esta forma es posible analizar los posibles efectos que la propagación de los errores puede tener en dos momentos diferentes del proceso de aplicación del cálculo de camino óptimo en el SIG. Para ello se aplicaron hasta tres tipos de análisis:

1. Un análisis de regresión por píxel entre las superficies de costes acumulados obtenidos y el mapa de costes acumulados original.
2. El cálculo de la Incertidumbre Media Cuadrática (IMC), entendida como la diferencia media entre el valor de coste acumulado de cada píxel de cada una de las simulaciones respecto al original.
3. La realización de una cartografía de los píxeles que forman parte de la ruta óptima un mayor número de veces, de todas las ejecuciones realizadas.

Del análisis de **regresión por píxel**, se deriva que el mapa de costes acumulado obtenido con las variables originales resulta bastante robusto en ambos casos. En el caso de el trazado de una nueva carretera entre Venturada y Colmenar Viejo, todos los coeficientes de determinación R^2 oscilan entre 0,98 y 0,99. Del análisis pormenorizado de cada una de las simulaciones, podríamos decir que son los casos en los que se ha introducido una distorsión espacial los que presentan los valores más bajos de correlación, por lo que podemos afirmar que podrían ser errores de este tipo los que más pudiesen influir en la variación de los resultados. No obstante hemos de reiterar que, con correlaciones superiores siempre al 0,98, no cabe duda de que los resultados son más que fiables. En el caso del cálculo de la trayectoria de la vía de ferrocarril entre Campo Real y Madrid, el resultado fue muy semejante, oscilando los coeficientes de determinación también entre 0,98 y 0,99.

Como hemos mencionado, se realizó un segundo análisis sobre los mapas de coste acumulado a partir del cálculo de la **Incertidumbre Media Cuadrática**, siguiendo la formulación de Gómez Delgado y Bosque Sendra (2004):

$$IMC = \sqrt{\frac{\sum_i (X_i - X_{it})^2}{n-1}}$$

siendo X_i = valor de coste acumulado de un píxel en cada superficie de coste acumulado obtenida a partir de la información distorsionada, X_{it} = valor de coste acumulado de un píxel en la superficie de coste acumulado obtenida a partir de la información original, n = número de ejecuciones.

En el caso 1, la mayor parte de los píxeles registraron una IMC baja, con una media de tan sólo 157 unidades de coste (siendo el máximo original de 30.000 unidades de coste acumulado), registrándose un porcentaje medio de variación de 2,4. Por otro lado, los valores máximos de IMC alcanzados suponen el 6,4% del valor máximo de coste acumulado original, lo que corrobora suficientemente la fiabilidad de nuestro resultado original. Como se puede derivar del mapa de distribución de la IMC (figura 9a), los valores máximos se concentran en los bordes de la imagen, justo donde se acumula el efecto de distorsión de la componente espacial de las distintas variables (ver ejemplo en figura 7). Si observamos el trazado de la ruta óptima original, comprobamos como toda esta parte del área de estudio es la que registra los valores de IMC más bajos.

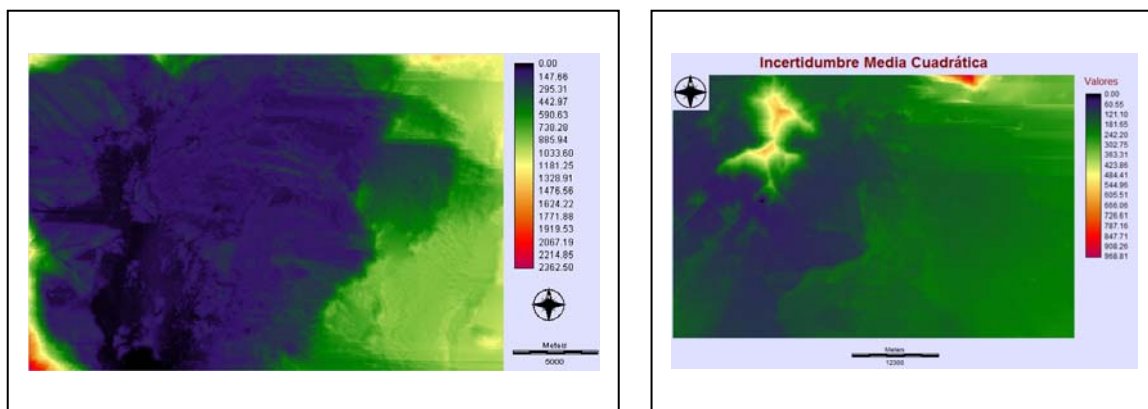


Figura 9. Incertidumbre Media Cuadrática del valor de coste acumulado en cada píxel. Izquierda (9a): nueva carretera entre Colmenar Viejo y Venturada; Derecha (9b): nueva vía férrea entre Madrid y Campo Real

En el caso 2 la situación es muy semejante. El valor medio de IMC registrado es de 302, aunque el valor máximo asciende, en este caso, hasta un 14% del valor máximo de coste acumulado original. Si observamos el mapa resultante en la figura 9b, comprobamos que aquí también se acumulan en el borde del mapa, concretamente en el norte del área de estudio. Por otro lado, los valores de IMC a lo largo del recorrido de la trayectoria obtenida originalmente vuelven a estar por debajo de la media o en torno a ella.

Por último, la **cartografía de los píxeles que más veces han sido seleccionados** para formar parte de la ruta óptima nos refuerza la confianza en los resultados obtenidos. A partir del simple análisis visual es posible sacar ya algunas conclusiones, pero también este sencillo análisis permite dar confianza concreta a cada píxel cuantificando el número de ensayos en los que aparece como solución. Así, en el caso 1 los píxeles que han sido seleccionados un mayor número de veces en el conjunto de los resultados con datos distorsionados aparecen, al menos, 12 veces en la trayectoria del camino original, alcanzando una media de 22,5 veces de un máximo de 32, lo que supone que la solución original aparece de media en un 73% de los ensayos. En la figura 10a encontramos la cartografía de las 32 rutas obtenidas, además de la distribución del número de veces que un píxel ha formado parte del camino de mínimo coste original. Si consideráramos tan sólo este último gráfico, la conclusión sería que el nivel de coincidencia es suficientemente elevado. Sin embargo, la cartografía conjunta de las 32 rutas nos permite determinar cómo es la distribución espacial de esas coincidencias. Así, comprobamos que el tramo medio de la trayectoria es bastante estable, sin embargo, en los tramos inicial y final de la ruta encontramos menos coincidencias, siendo el recorrido de algunas trayectorias bastante diferentes. En el caso 2, los resultados vuelven a ser también positivos, encontrando bastantes tramos seleccionados más de 15 veces, de las 25 simulaciones ejecutadas (figura 10b).

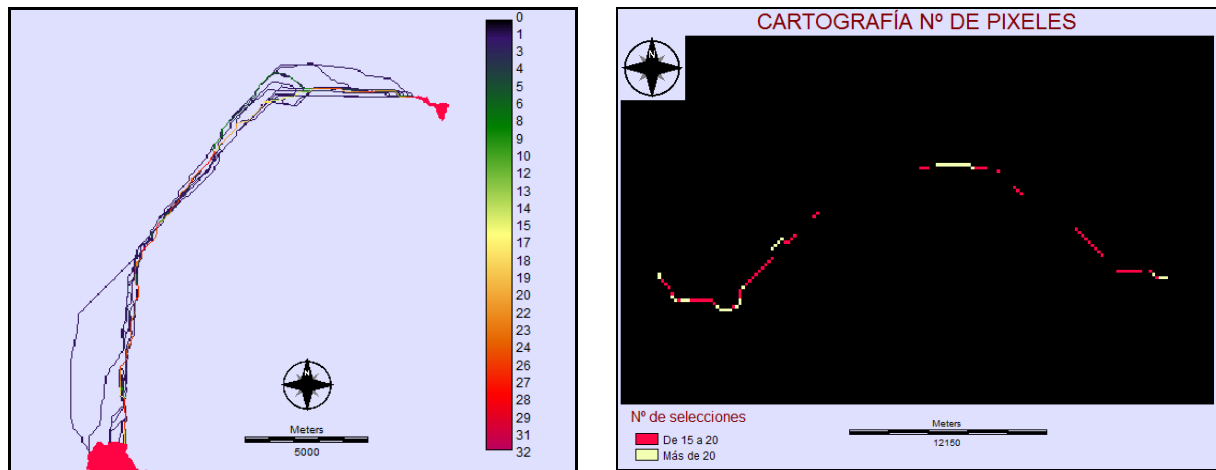


Figura 10. Cartografía de los píxeles que más veces han sido seleccionados para formar parte de la trayectoria. Izquierda (10a): nueva carretera entre Colmenar Viejo y Venturada. Derecha (10b): nueva vía férrea entre Madrid y Campo Real

5. DISCUSIÓN Y CONCLUSIONES

Ante la incertidumbre que presenta la utilización de toda información geográfica en cuanto a su calidad, y la dificultad inherente a descubrir sus causas, el análisis de incertidumbre se presenta como una metodología sencilla que permite cuantificar la confianza que merecen los resultados obtenidos por un determinado modelo, sin tener que realizar un esfuerzo por conocer los posibles errores existentes en la base de datos de partida.

Por otro lado, la aplicación de un análisis de sensibilidad de los resultados ante pequeñas modificaciones introducidas en los parámetros de entrada del modelo, resulta adecuada para comprobar la estabilidad de los resultados del modelo. Así, la utilización conjunta de ambos análisis como parte del proceso de validación ha resultado efectivo, flexible y fácilmente aplicable en el cálculo de camino óptimo, permitiendo controlar el riesgo que se asumiría al tomar cualquier decisión basándonos en los resultados de este modelo. Sustituir un proceso de verificación estricto por uno de validación más flexible puede animar a los investigadores a realizar este esfuerzo una vez concluido el proceso de análisis propiamente dicho y así, asegurar una cierta confianza en los resultados.

El elevado coeficiente de correlación entre los resultados obtenidos al calcular el camino de mínimo coste con las variables de entrada originales y al hacerlo con las distorsionadas, la baja incertidumbre media cuadrática (IMC) de cada píxel con respecto al valor de coste acumulado original, y la utilización del mayor número de veces que aparece un píxel en los resultados ofrecidos por el modelo, permiten validar el cálculo de camino de mínimo coste sin necesidad de conocer la fuente exacta del error en la información de partida. Además, estas técnicas de análisis propuestas resultan de sencilla aplicación con las herramientas convencionales que incorpora cualquier SIG.

La imposibilidad de utilizar en este caso análisis de tipo global como Sobol' o FAST, debido a la dificultad que supone simular la ejecución del modelo espacial en un programa que no tiene en cuenta estas características como SIMLAB y la imposibilidad de ejecutarlo fuera del programa debido a la carga computacional necesaria para hacerlo, nos ha obligado a utilizar otros análisis más sencillos para comprobar la estabilidad de los resultados. Por otro lado, no habiendo podido averiguar cuáles y en qué cantidad influyen los factores del modelo en la variación de los resultados, ha sido inviable utilizar los resultados del análisis de sensibilidad para simplificar el modelo (reteniendo tan solo aquellos factores más influyentes).

6. REFERENCIAS BIBLIOGRÁFICAS

- Backhoff Pohls, M. A. (2005): *Transporte y espacio geográfico*. México, UNAM.
- Burnicki, A.C.; Brown, D.G.; Goovaerts, P. (2007): Simulating error propagation in land-cover change analysis: the implications of temporal dependence. *Computers, Environment and Urban Systems*, 31 (3): 282-302.
- Crosetto, M. y Tarantola, S. (2001): Uncertainty and sensitivity analysis: tools for GIS-based model implementation. *International Journal of Geographical Information Science*, 15 (5): 415-437.
- Crosetto, M.; Moreno Ruiz, J.A. y Crippa, B. (2001): Uncertainty propagation in models driven by remotely sensed data. *Remote Sensing of Environment*, 76 (3): 373-437.
- Dane, G.Z. y Tecim, V. (2007): GIS based route determination for light rail systems: A case study in Izmir, Turkey. *Local governance and sustainable development. Proceedings of the Joint Congress of the European Regional Science Association (47th) and ASRDLF (44th)*, Paris, August 29, pp. 1-14.
- Delavar, M.R. y Naghibi, F. (2003): Pipeline routing using Geospatial Information System Analysis, *Proceedings of ScanGIS'2003*, Espoo, Finland, June 4-6.

- Feldman, S.; Pelletier, R.; Walser, E.; Smoot, J.; Douglas, A., (1995): A prototype for pipeline routing using remotely sensed data and geographic information system analysis. *Remote Sensing of Environment*, 53: 123-131.
- Ganskopp, D.; Cruz, R.; Johnson, D., (2000): Least-effort pathways?: a GIS analysis of livestock trails in rugged terrain. *Applied Animal Behaviour Science*, 68 (2): 179-190.
- Gómez Delgado, M. (2003): *SIG y toma de decisiones: control del riesgo a partir de análisis de sensibilidad y análisis de incertidumbre*. Tesis doctoral. Universidad de Alcalá de Henares.
- Gómez Delgado, M.; Rodríguez Espinosa, V.R.; Rodríguez Durán, A.E.; Chuvieco Salinero, J.; Chuvieco Salinero, E. (1995): Diseño de carreteras mediante un sistema de información geográfica: Costes de construcción y costes ambientales. *Ciudad y Territorio (Artículos)*, 104: 361-376.
- Gómez Delgado, M. y Bosque Sendra, J. (2004): Aplicación de análisis de incertidumbre como método de validación y control de riesgo en la toma de decisiones. *GeoFocus (Artículos)*, 4: 179-208.
- Gómez Delgado, M., Barredo Cano, J., (2005): *Sistemas de información geográfica y evaluación multicriterio en la ordenación del territorio*. Paracuellos de Jarama, Editorial RA-MA.
- Gómez Delgado, M. y Tarantola, S. (2006): Global sensitivity analysis, GIS and multi-criteria evaluation for a sustainable planning of a hazardous waste disposal site in Spain. *International Journal of Geographical Information Science*, 20 (4): 449-466.
- Gutiérrez Puebla, J., García Palomares, J. C., (2005): Cambios en la movilidad en el área metropolitana de Madrid: el creciente uso del transporte privado. *Anales de Geografía de la Universidad Complutense*, 25: 331-351.
- Haining, R.; Giuseppe, A.; Griffith, D. (1998): Error propagation modelling in raster GIS: overlay operations. *International Journal of Geographical Information Science*, 12 (2): 145-167.
- Heuvelink, G.B.M. (1998): *Error propagation in environmental modelling with GIS*. London, Taylor & Francis.
- Hunter, G.J. (1999): Managing uncertainty. En P.A Longley; M.F. Goodchild; D.J. Maguire y D.W. Rhind (Eds.): *Geographical Information Systems*. Volume II, New York, John Wiley & Sons, Inc., pp. 633-641.
- López Romero, R. (2005): Calculo de Rutas Óptimas mediante SIG en el territorio de la ciudad celtibérica de Segeda. *Propuesta Metodológica. SALDVIE*, 5: 95-111.
- Rae, C.; Rothley, K. and Dragicevic, S. (2007): Implications of error and uncertainty for an environmental planning scenario: a sensitivity analysis of GIS-based variables in a reserve design exercise, *Landscape and urban planning*. 79: 210-217.

- Rykiel, E.J. Jr. (1996): Testing ecological models: the meaning of validation, *Ecological modelling*, 90: 229-244.
- Santiago, E. de (2008): Madrid "ciudad única" (II). La explosión urbana en la región madrileña y sus efectos colaterales". *URBAN*, 13: 138-164
- Seguí Pons, J.M. y Martínez Reynés, M.R.(2003): Pluralidad de métodos y renovación conceptual en la geografía de los transportes del siglo XXI. *Scripta Nova. Revista Electrónica de Geografía y Ciencias Sociales*, VII (139).
- Yeh, A.G. and Li, X. (2006): Errors and uncertainties in urban cellular automata. *Computers, Environment and Urban Systems*, 30: 10-28.
- Yildirim, V., Nisanci, R. and Reis, S. (2006): A GIS based route determination in linear engineering structures information management (LESIM). *Shaping the Change, Proceedings of XXIII FIG Congress*, Munich, Germany, October 8-13.