

## ESTIMACIÓN DE COSTES DE DESARROLLO DE PROYECTOS DE SISTEMAS DE INFORMACIÓN GEOGRÁFICA

José Luis González Vázquez

*Aqualia Infraestructuras*

José Manuel Mesa Fernández, Ramiro Concepción Suárez

*Universidad de Oviedo*

María Ángela Alonso Fernandez

*Seresco*

### Resumen

En la actualidad el empleo de los Sistemas de Información Geográfica (SIG) se ha extendido enormemente. La gran variedad de aplicaciones, desde el ámbito más local hasta la cartografía más extensa, los distintos tipos y configuraciones del hardware and software empleados, la incorporación de aspectos espacio-temporales o las herramientas de análisis y explotación de la información utilizadas, dan lugar a proyectos SIG de características muy diferentes. Este trabajo tiene como objetivo proponer un modelo general de estimación del esfuerzo necesario para la realización de los proyectos SIG. Para ello se identifican inicialmente el conjunto de factores que permitan caracterizar un proyecto SIG. Posteriormente se evalúa la aplicabilidad de distintas técnicas de estimación de esfuerzo y su adecuación a este problema. Finalmente la validación del modelo de estimación propuesto se realiza a partir de la aplicación a distintos casos particulares.

**Palabras clave:** *Modelos de estimación; Sistemas de Información Geográfica, Esfuerzo de desarrollo*

### Abstract

At present, the use of Geographic Information Systems (GIS) has expanded enormously. The wide variety of applications, from the most local to the most extensive mapping, the different types and configurations of hardware and software used, the incorporation of spatial and temporal aspects or tools of analysis and information exploitation, result in GIS projects with very different characteristics. This paper aim is to propose a general model to estimate the effort required to carry out GIS projects. Initially a set of factors to characterize a GIS project are identified. Afterwards, the applicability of different effort estimation techniques and their adaptation to this problem is evaluated. Finally, the validation of the proposed estimation model is drawn by its application to different cases.

**Keywords:** *Management models, Geographic Information Systems, Effort Estimation*

## 1. Introducción

Una de las constantes en los manuales escritos sobre SIG es la enorme complejidad en su definición. La causa fundamental de esta dificultad es que integran en un mismo concepto:

- Datos y teorías sobre los datos.
- Hardware y software.
- Diversos intereses (científicos, de gestión y comerciales).
- Diversas disciplinas científicas (matemáticas, informática, cartografía, geografía, biología, ingeniería, etc.) que previamente aparecían aisladas.

**Figura 1: Componentes de un Sistema de Información Geográfica**



En general, un Sistema de Información consiste en la unión de información y herramientas informáticas (programas) para su análisis con unos objetivos concretos. En el caso de los SIG, se asume que la información incluye la posición en el espacio de los datos.

La base de un Sistema de Información Geográfica es, por tanto, una serie de capas de información espacial en formato digital que representan diversas variables (formato raster), o bien capas que representan objetos (formato vectorial) a los que corresponden varias entradas en una base de datos enlazada. Esta estructura permite combinar, en un mismo sistema, información con orígenes y formatos muy diversos incrementando la complejidad del sistema.

Los Sistemas de Información Geográfica se han desarrollado a partir de la unión de diversos tipos de aplicaciones informáticas: la cartografía automática tradicional, los sistemas de gestión de bases de datos, las herramientas de análisis digital de imágenes, los sistemas de ayuda a la toma de decisiones y las técnicas de modelización física. Se ha llegado a considerar a los SIG como un enlace entre la Geografía y la Informática al igual que la Geomorfología enlazaría Geografía y Geología.

Una de las primeras percepciones que se tienen de un SIG son las salidas gráficas a todo color, impresas o en la pantalla de un ordenador. Conviene recordar sin embargo que hay una diferencia fundamental entre los programas de manejo de gráficos y los SIG. En los primeros, lo fundamental es la imagen que vemos, siendo irrelevante como se codifique, en un SIG la imagen es sólo una salida gráfica sin mayor importancia, lo relevante son los datos que se están representando.

Se han dado diversas definiciones de Sistema de Información Geográfica, las siguientes que constituyen ejemplos claros de los modos habituales de concebir un SIG:

- La primera referencia al término SIG aparece en Tomlinson (1967) referida a una aplicación informática cuyo objetivo era desarrollar un conjunto de tareas con información geográfica digitalizada. Se trataba del Sistema de Información Geográfica de Canada (CGIS)
- Para Berry (1987) un Sistema de Información Geográfica es un sistema informático diseñado para el manejo, análisis y cartografía de información espacial.
- Para Burrough (1988), se trata de un conjunto de herramientas para reunir, introducir, almacenar, recuperar, transformar y cartografiar datos espaciales sobre el mundo real para un conjunto particular de objetivos. El National Center for Geographic Information and Analysis de USA los define como Sistema de hardware, software y procedimientos elaborados para facilitar la obtención gestión, manipulación, análisis, modelado, representación y salida de datos espacialmente referenciados para resolver problemas complejos de planificación y gestión (NC- GIA, 1990).

Como se observa, estas definiciones no sólo son consecutivas en el tiempo, sino que además cada una supone un mayor nivel de complejidad sobre la anterior. La primera hace referencia únicamente a las bases de datos espaciales; la segunda y la tercera a las herramientas (software) de tratamiento de estos datos, el típico paquete de módulos de SIG (GRASS, Erdas, ArcInfo, Idrisi, Atlas-Gis, etc.); finalmente la tercera incluye el hardware utilizado y los procedimientos complementarios que puedan ser necesarios.

Aunque en sentido estricto no sería necesario, se han desarrollado un tipo específico de aplicaciones informáticas para el manejo de estos sistemas. Estos programas es lo que popularmente (y equivocadamente) se conoce cómo SIG (IDRISI, ArcInfo, GvSig, etc.), pero que realmente constituyen tan sólo un componente de lo que es realmente el sistema. Podríamos considerar, en sentido amplio que un SIG está constituido por:

1. Bases de datos espaciales en las que la realidad se codifica mediante unos modelos de datos específicos.
2. Bases de datos temáticas cuya vinculación con la base de datos cartográfica permite asignar a cada punto, línea o área del territorio unos valores temáticos incluidos en los registros de las tablas.
3. Conjunto de herramientas que permiten manejar estas bases de datos de forma útil para diversos propósitos de investigación, docencia o gestión (Software).
4. Conjunto de ordenadores y periféricos de entrada y salida que constituyen el soporte físico del SIG (Hardware). Estas incluyen tanto el programa de gestión de SIG cómo otros programas de apoyo.
5. Comunidad de usuarios que pueda demandar información espacial, cada vez más importante por el amplio desarrollo de sistemas de esta índole en Internet en los últimos años.
6. Administradores del sistema encargados de resolver los requerimientos de los usuarios bien utilizando las herramientas disponibles o bien produciendo nuevas herramientas.

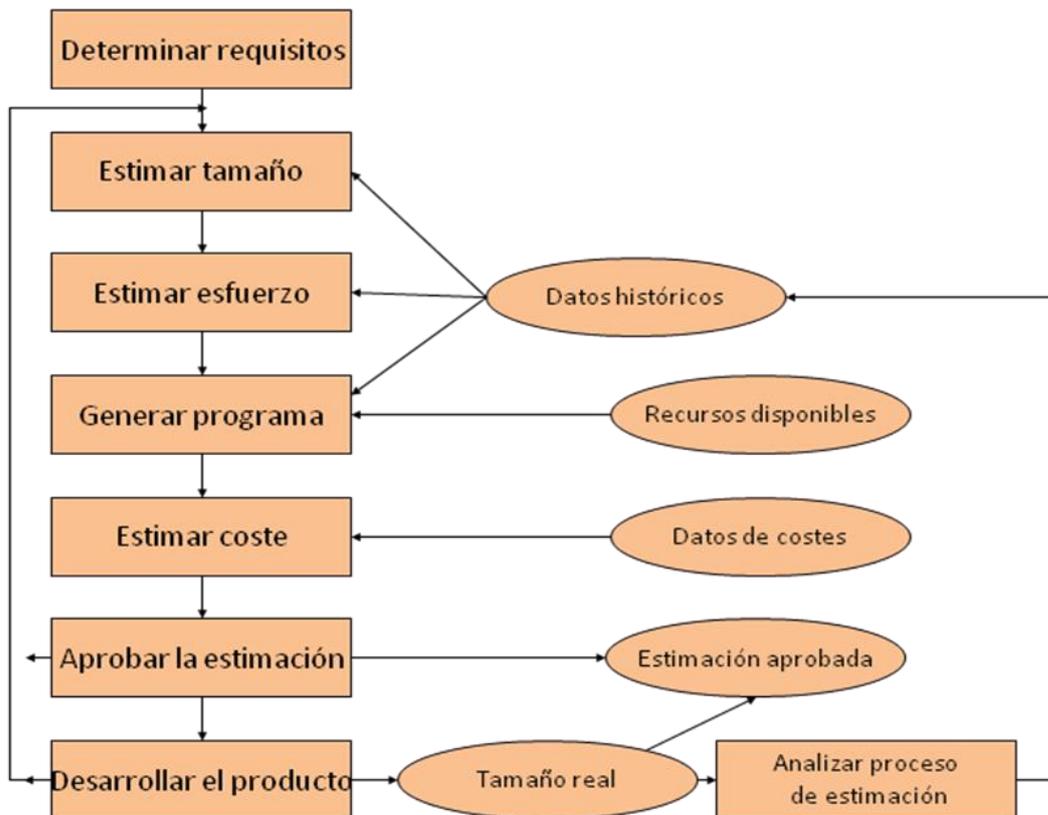
## 2. El problema de la estimación de esfuerzos en SIG

A la hora de enfrentarse al desarrollo de un Sistema de Información Geográfica, es necesario considerar, como para todo proyecto, una planificación adecuada del mismo que permita conocer sus necesidades de recursos, coste y duración.

Un modelo de estimación del esfuerzo debe poseer las siguientes características (Varas, 1995):

1. Capacidad de adaptación a la productividad de la organización.
2. Considerar el coste de comunicación entre los involucrados.
3. Incorporar guías útiles para estimar aquellos parámetros que son subjetivos o no se deducen en forma explícita a partir del modelo.
4. Debe ser usable.
5. Debe reflejar etapas simples de entender y definidas en forma precisa.
6. Objetividad.
7. Coste efectivo.
8. Capacidad de proveer medios para adaptar los cambios en el desarrollo.
9. Debe permitir una estimación temprana.

Figura 2: El proceso de estimación de esfuerzos.



Habitualmente los procesos de estimación siguen una serie de etapas similares a las que aparecen reflejadas en la Figura 2.

La mayoría de los métodos incluyen modelos empíricos de estimación y poseen como variable definitoria del coste principal el tamaño de la aplicación a desarrollar, lo que es suficientemente difícil de estimar en los SIG, e implica la necesidad de encontrar un modelo adaptado a esta tarea que permita la obtención de un método más fácil de usar.

Para el caso de los Sistemas de Información Geográfica, se puede observar que en la actualidad las herramientas de desarrollo proveen la capacidad de disminuir substancialmente el esfuerzo de desarrollo de herramientas gestoras de bases de datos, centrando las necesidades en la adquisición, análisis y adecuación de la información. (Moles, 1995)

En la mayoría de los proyectos SIG que se desarrollan actualmente se utilizan técnicas de gestión que apoyan débilmente a la planificación, consistente generalmente en la utilización de software comercial que facilita la carga de datos espaciales y geográficos. Estos productos parten de la base de que los recursos a consumir son conocidos a priori, aspecto que no siempre se cumple.

No resulta factible en este tipo de proyectos la descomposición en unidades de obra o la posibilidad de acudir a una base de datos de precios comercial al igual que sucede, por ejemplo en la construcción. Si acaso, sí se pueden referenciar determinados costes de datos espaciales o cartográficos elaborados con antelación y de forma más o menos genérica, como los productos de los institutos cartográficos u otros organismos análogos. Sin embargo las necesidades cartográficas específicas o la adecuación y tratamiento de datos espaciales relativizan en gran parte la importancia de estos catálogos.

Por otro lado, dada la heterogeneidad de los datos, que a su vez presentan diversos tamaños y características, así como la multitud de tecnologías disponibles y la amplitud de necesidades que deben cubrir este tipo de proyectos, se ha hecho tradicionalmente difícil la planificación de costes relativos a su implementación. Resulta, pues, necesario establecer un modelo de estimación de esfuerzos adecuado que intente cubrir esas necesidades de la forma más completa posible.

Todos los puntos mencionados anteriormente, dificultan que la utilización de modelos de gestión sea una práctica generalizada en los administradores de proyectos de desarrollo de SIG.

### 3. Desarrollo del modelo

Con el fin de abordar el problema objeto del estudio debe configurarse una metodología de estimación general que describa las componentes y caracterización de este tipo de desarrollos, con el fin de traducir o convertir dichas características en indicadores de esfuerzo.

Las bases que se plantean, deben permitir constituir una herramienta de consultoría dirigida al análisis, diseño y costes de desarrollo que faciliten el conocimiento de la implementación tanto a nivel operativo como estructural, y cuya clave principal debe ser la versatilidad en su adecuación al desarrollo de cualquier proyecto de Sistemas de Información Geográfica, independientemente del volumen del mismo.

Teniendo en cuenta las etapas habituales en la estimación de costes de desarrollo y adaptándolas al caso de los Sistemas de Información Geográfica, se han definido las siguientes fases de desarrollo del estudio:

- **Fase 1 Caracterización:** Supone la definición de un entorno, es decir, el conocimiento de las necesidades y características de la organización y los usuarios del SIG. Dichas necesidades se traducirán en un conjunto de propiedades (cliente/servidor, multiusuario, etc.) y una serie de productos a obtener (consultas, informes, mapas...) que permiten

caracterizar progresivamente el proyecto, por ejemplo, seleccionando la tecnología más adecuada (hardware, software) para cada caso particular.

El punto de partida de esta definición será de este modo un marco general de características y componentes de los sistemas SIG que debe estar convenientemente actualizado para incorporar las nuevas posibilidades técnicas que se van incorporando en este ámbito. Teniendo en cuenta este esquema se realizarán entrevistas en profundidad con los principales usuarios del futuro sistema. Los resultados de esta primera etapa supondrán la selección de:

- El SGBD, el tamaño y otras características de la base de datos.
- La información geográfica necesaria, el tipo (vectorial, raster), el proceso de captura, la inserción de datos, etc.
- Los requisitos del software SIG a utilizar o a desarrollar.
- Las especificaciones del hardware necesario.

Este primer paso define las características del sistema, abordando el estudio de las particularidades de este tipo de proyectos a partir de los conceptos generales y comunes que los forman, profundizando en sus propiedades hasta la definición de una serie de componentes representativos que, a modo de indicadores, posibiliten en fases posteriores la aplicación de diversas técnicas de estimación.

- **Fase 2 Estimación:** Selección de las técnicas de estimación más adecuadas para cada uno de los componentes identificados y caracterizados en la fase anterior.

La estimación de esfuerzos de desarrollo de software se define tradicionalmente por la aplicación de técnicas orientadas a la cuantificación de la complejidad del sistema (puntos de función, casos de uso, etc.), y su estudio constituye un apartado necesario en la evaluación de las componentes correspondientes de cualquier proyecto SIG.

En el caso de la adquisición, tratamiento e inserción de datos, tanto gráficos como alfanuméricos, el modelo debe definir una serie de parámetros de esfuerzo cuya cuantificación y homogeneización sea factible a una determinada escala. Se propondrán fórmulas dependientes de la mano de obra, medios disponibles y las características propias del medio geográfico estudiado, entre otros factores, intentando cubrir las opciones que se presentan habitualmente en estas fases del proyecto.

Posteriormente se realizará una aplicación de las técnicas, generando una estimación de esfuerzo (horas de trabajo) o bien unos indicadores indirectos que luego se puedan traducir a unidades de tiempo y posteriormente de coste.

- **Fase 3 Integración:** Consistirá en la unificación de los resultados obtenidos para cada componente en la fase anterior y que permitirá generar una estimación global para el proyecto. Se requerirán nuevos parámetros de ajuste en función de distintos aspectos como plazos, disponibilidad de recursos, etc. que no se hubieran introducido en fases anteriores.

La principal dificultad de esta fase estriba en la necesidad de homogeneizar las estimaciones obtenidas para cada componente de la fase anterior, así como considerar las distintas interacciones entre los mismos.

**Figura 3: Fases metodológicas.**



A continuación se describe en mayor profundidad el contenido de cada una de las tres etapas en las que se ha estructurado la metodología o modelo de estimación.

### 3.1. Fase de caracterización

Anteriormente se han definido los componentes básicos de los SIG, sin embargo, estos componentes son transversales al proyecto y no aportan *per se* información cuantificable que permita estimar de modo fehaciente el coste de implementación del mismo. Resulta, pues, necesario analizar dichos componentes de forma jerárquica, realizando una caracterización que posibilite sugerir indicadores específicos para cada apartado. Dicha caracterización tiene la finalidad de incorporar a cada uno de esos elementos una serie de parámetros de productividad sobre los que realizar una estimación de su esfuerzo que finalmente sería extrapolable a la totalidad del desarrollo del proyecto.

#### 3.1.1. Base de datos

Una base de datos (BD) es un conjunto de información digital relacionada. En el ámbito del presente estudio las BD contienen información geográfica, relativa a posiciones o elementos espaciales, en incluso como almacén de información de la propia cartografía base del SIG.

Uno de los sistemas más utilizados son las bases de datos relacionales, esto es, sistemas cuya información se almacena mediante tablas, en filas y columnas ordenadas denominadas 'registros' y 'atributos' o 'campos' respectivamente.

En el proceso de construcción de la BD no existe un único procedimiento genérico que se ajuste para cualquier diseño, se puede hablar más bien de una pauta a seguir. El proceso de planificar una base de datos no es lineal, existen bucles retroalimentativos, especialmente entre el estudio piloto, los diseños conceptuales y lógicos y el diseño físico.

### 3.1.1.1 Tipo de SGBD

Un SGBD (sistema gestor de base de datos) es el software específico, encargado de servir de interfaz entre la BD, el usuario y las aplicaciones que la utilizan. El propósito general de los sistemas de gestión de bases de datos es el de gestionar de manera clara, sencilla y ordenada un conjunto de datos que posteriormente se convertirán en información relevante para una organización.

De este modo el SGBD poseerá siempre determinadas y diversas estructuras de almacenamiento de datos con soporte para la definición de restricciones de integridad, mecanismos para la gestión de los mismos y subsistemas de control de la concurrencia, acceso y fallos para impedir que se produzcan inconsistencias ante accesos simultáneos o no autorizados o pérdida de datos.

Se puede clasificar los SGBD en cuanto a su estructura interna de funcionamiento, que será dual o integrada.

En la estructura dual el sistema separa la organización y la gestión de la información temática y la espacial, de modo que opera un software específico sobre cada componente. Este modelo es habitual en un Sistema de Información Geográfica vectorial, donde se utiliza el modelo relacional para la información temática y otro de base topológica orientado a objetos (descripción de las coordenadas de los objetos gráficos y de su posicionamiento relativo) para la base de datos espacial.

Por otra parte, los sistemas integrados gestionan únicamente una base de datos relacional para todo tipo de información del SIG. Esto conlleva ciertos inconvenientes, dado que la estructura de datos espaciales se presta poco a ser gestionada exclusivamente mediante bases de datos relacionales, ya que las tablas implican una dimensión fija que los registros que no siempre es adecuada a elementos geográficos que suelen tener longitudes de puntos extraordinariamente variables.

El modelo dual, por el contrario, resulta más rápido frente al integrado, pero adolece de falta de claridad en la organización interna de los datos. Al existir dos sistemas de gestión de bases de datos pueden presentarse también problemas determinados de ensamblaje que pueden afectar a la integridad del sistema. Por su parte el modelo integrado reduce este tipo de riesgos al aportar una mayor robustez.

Dos de los principales proveedores de SGBD comerciales han lanzado extensiones espaciales de base de datos orientadas a objetos: IBM ofrece dos soluciones - DB2 Spatial Extender e Informix Spatial Datablade - y Oracle tiene una opción espacial -Oracle Spatial -. Aunque existen diferencias en la tecnología, el alcance y las capacidades de estos sistemas, todos ellos proporcionan funciones básicas para almacenar, gestionar y consultar los objetos geográficos. Ninguno de ellos es un sistema completo de software de SIG en sí mismo. El objetivo de estas extensiones es el almacenamiento de datos, recuperación y gestión, y no tienen capacidad real para la edición geográfica, la cartografía y el análisis. En consecuencia, deben ser utilizados conjuntamente con un SIG excepto en el caso de las aplicaciones más sencillas centradas en consultas.

### 3.1.1.2. Tamaño BD

El tamaño de la Base de Datos estará siempre ligado al tipo y cantidad de datos geográficos que serán gestionados por la misma.

De este modo, para determinar el tamaño de la base de datos no basta con atenerse exclusivamente a las fases del desarrollo de la misma, es necesario conceptualizar una serie de indicadores que permitan estimar el esfuerzo empleado en dicha tarea. Así pues se

debe profundizar en sus componentes para llegar a un nivel mínimo que facilite una cierta cuantificación de magnitudes de esfuerzo:

- **Tablas:** Representa un tipo de modelado de información, donde se guardan los datos recogidos por un programa. Independientemente de su carácter espacial, se dividen en dos tipos básicos en función con la información recogida:
  - **Tipo objeto:** la tabla representa algo tangible como un conjunto de personas, conjunto de lugares o conjunto de cosas. Además, todos los objetos, comparten las mismas características las cuales pueden ser almacenadas como datos.
  - **Tipo evento:** la tabla representa algo que ocurre en un momento dado y tiene características que se desean registrar en la misma línea que en el caso de los tipos de objeto.

Su caracterización viene dada por los siguientes elementos:

- Número de campos
- Tipo de campos: Es la estructura más pequeña representa una propiedad o característica de un objeto/evento de la tabla.
- Número de registros: un registro representa a uno de los objetos/eventos de la tabla. No importa el orden de las filas.
- **Índices:** Clave primaria (PK, primary key) de una tabla. Es un atributo simple (clave simple) o un grupo de atributos (clave compuesta) que identifica de manera única cada registro de la tabla. Es obligatorio que cada tabla tenga una PK para asegurar su integridad.

Las principales dificultades que se encuentran en este apartado residen en la posibilidad de encontrarnos ante escenarios no definidos, especialmente en la etapa inicial del diseño, que dificulten la definición de los indicadores anteriormente citados, por lo que puede ser necesario establecer una escala general en función del volumen de datos previsto que nos dé un orden de magnitud.

### 3.1.1.3. Inserción/carga de datos

Una vez solventado las cuestiones de diseño, y suponiendo accesible toda fuente de datos, es necesario considerar el tiempo invertido en la carga de los mismos en la base de datos. En este apartado el método de inserción o carga será determinante para la estimación de esfuerzos, dado que la elección de una u otra metodología puede suponer considerables variaciones en los costes del proyecto.

Métodos de inserción o carga de información en la base de datos:

- **Manual:** La inserción manual de datos debe considerarse únicamente cuando los medios al alcance no permitan su automatización, el tamaño de los mismos sea aceptable y su complejidad relativa para evitar cometer errores que deberán ser subsanados posteriormente y comprometerían la integridad de la base de datos. De modo que en función de la cantidad de datos y una estimación de la velocidad de inserción se podría estimar el coste de inserción.
- **Automatizada:** Lo más común es que la importación de datos sea automatizada desde:
  - Fichero
  - Otras BD

En este caso la labor la realiza un software, casi siempre el SGBD o bien un desarrollo adecuado la tipología y tamaño de la información a adquirir. Para cuantificar esto se desarrollarán métodos de importación.

### **3.1.2. Adquisición de información geográfica**

En los inicios de los SIG, cuando los datos geográficos aún eran bastante escasos, la recolección de información era la tarea principal del proyecto y por lo general consumía la mayor parte de los recursos disponibles. Incluso hoy en día la recogida de datos sigue siendo un proceso largo, tedioso y caro que normalmente representa el 15-50% del coste total de un proyecto SIG. Los costes de captura de datos puede ser de hecho mucho más significativos porque en muchas organizaciones (especialmente los que son financiados por la administración) los gastos de personal se suponen fijos y no se utilizan en la contabilidad presupuestaria. Si se excluyen los gastos de personal del presupuesto de un proyecto SIG los gastos relativos a la adquisición de datos puede conformar de hasta 60-85% del coste final, se trata, por tanto de un apartado de importancia en el desarrollo del estudio (Longley, 2005).

En todos los proyectos, la recopilación de datos implica una serie de etapas secuenciales. El flujo de trabajo comienza con la planificación, seguido por la preparación, digitalización / transferencia, edición y mejora y, por último, la evaluación.

La planificación a la hora de la recopilación de datos resulta imprescindible, pues de ella va a depender en gran medida el desarrollo restante del proyecto el proceso puede conllevar múltiples tareas, tales como la obtención de datos, recopilación de fuentes de cartografía, filtrado de elementos de mala calidad, edición de imágenes digitalizadas de mapas, la eliminación de ruido (datos no deseados, tales como manchas en una imagen digitalizada del mapa). También puede implicar el establecimiento de sistemas de hardware y software automatizados para aceptar datos. La digitalización y transferencia son las etapas donde se realiza la mayor parte del esfuerzo realizado. La captura de datos comprende en realidad muchos más procesos que la mera digitalización y anteceden a la edición y la adecuación de los datos, que abarcan abarca diversas técnicas de validación, para corregir errores y mejorar la calidad. Seguidamente se puede proceder a un proceso de evaluación cualitativo o cuantitativo, con el fin de identificar los problemas surgidos durante la adquisición de los datos. Dado que todos los grandes proyectos de datos implican múltiples etapas, este flujo de trabajo iterativo con fases anteriores puede contribuir a mejorar las partes posteriores del proyecto final.

#### **3.1.2.1. Captura directa de datos**

Este tipo de captura de información consiste en la medición directa de los objetos sobre el terreno. Las mediciones de los datos digitales pueden entrar directamente en la base de datos del SIG, o bien pueden residir en un archivo temporal para ser procesados antes de su inserción. Aunque la primera opción resulta preferible, ya que minimiza la cantidad de tiempo y la posibilidad de errores, no siempre es posible debido a la logística y a que la coordinación directa entre los dispositivos de recopilación de datos y bases de datos SIG no siempre es factible.

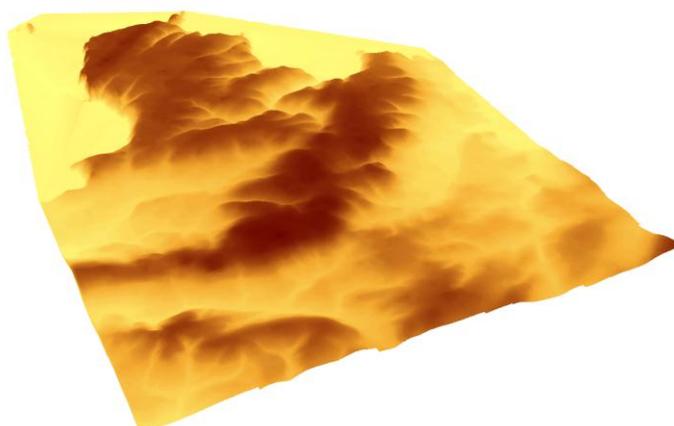
#### **3.1.2.2. Captura directa de datos raster**

Las fuentes principales de captura de datos raster son la teledetección y la fotogrametría. En términos generales, la teledetección es una técnica utilizada para obtener información acerca de las propiedades físicas, químicas y biológicas de objetos sin toma de contacto directo con los mismos. La información se deriva de las mediciones de la cantidad de radiación electromagnética reflejada, emitida o dispersada por los objetos. Una serie de

sensores, que operan en todo el espectro electromagnético que va desde las microondas a longitudes de onda visibles, capturan los datos de forma remota (Maguire, 1995).

Los sensores pasivos dependen de la radiación solar reflejada o emitida la radiación terrestre, mientras que los sensores activos generan su propia fuente de radiación electromagnética. Las plataformas desde las que operan estos instrumentos son igualmente diversas, aunque los satélites que orbitan la Tierra y los aviones son con diferencia los más comunes, también se emplean helicópteros, globos, mástiles, etc. Desde el punto de vista de los SIG, la resolución es la característica clave sobre la que se definen todos los sistemas que operan mediante teledetección. Hay tres componentes de la resolución: espacial, espectral y temporal, cuyos costes deben considerarse por separado y pueden constituir en sí mismos unos indicadores ampliamente decisivos para cuantificar los costes de los mismos.

**Figura 4: Modelo de datos raster obtenido por técnicas de captura directa**



Todos los sensores, así como las propiedades espaciales, espectrales y temporales de las imágenes deben supeditarse a las necesidades de almacenamiento y tratamiento de los datos, así como al ancho de banda disponible.

La misión del vuelo fotogramétrico tiene por objeto, el sobrevolar la zona a altura y velocidad constante, describiendo una serie de trayectorias (pasadas), paralelas entre sí, mediante su control de deriva, tomando fotografías de la zona de estudio. Dichas imágenes deben ser procesadas posteriormente para corregir defectos de perspectiva y proceder a su georreferenciación.

En fotogrametría deben considerarse la extensión y orografía del terreno, la escala del modelo, las especificaciones técnicas del equipo de toma de imágenes (avión, cámara, focal, recubrimiento) y la exactitud y finalidad de las tomas.

### **3.1.2.3. Captura directa de datos vectoriales**

La toma directa de datos vectoriales en campo resulta crucial para la particularización de la información espacial en pequeñas extensiones de terreno que deben ser representadas a gran escala. Estos datos pueden tomarse de forma única, o complementando y particularizando otros provenientes de otras fuentes. En ocasiones puede ser necesario realizar un levantamiento topográfico para corregir errores, indeterminaciones o inexactitudes debidas a la componente temporal de la cartografía disponible.

Tradicionalmente se ha recurrido a técnicas topográficas que consisten en la medición de ángulos y distancias sobre el terreno, lo cual permite definir una serie de puntos mediante coordenadas 3D que constituirán una representación gráfica de los elementos reales.

Para estimar el esfuerzo de esta fase es necesario considerar la escala, la orografía, el número de puntos a definir, así como el tipo de elementos (capas) que se van a medir sobre el terreno y los equipos y personal destinados a tal fin.

#### 3.1.2.4. Captura indirecta de datos

Se denomina así a la obtención de datos geográficos provenientes de otras fuentes, como mapas en papel, fotografía aérea y otros documentos. Para la obtención de datos raster se recurre al escaneado de la cartografía impresa, mientras que para la definición de los datos vectoriales es necesario proceder a la digitalización sobre tableta o pantalla.

#### 3.1.2.5. Captura de datos raster mediante escáneres

Cada vez menos común por la existencia de numerosa información digital, aún se utiliza en determinados proyectos que parten de datos antiguos o mapas cuya información resulta específica y no ha sido informatizada convenientemente.

Para cuantificar el esfuerzo de la captura de datos mediante escaneado deben considerarse por separado la entrada y la salida digital de los mismos:

- **INPUT:** La cuantía de los datos en papel resulta decisiva, en tanto que mantendrá una relación directamente proporcional con la cantidad de esfuerzo empleado. Debe consignarse el formato y dimensiones de los originales, o bien una superficie de papel total que debe ser procesada. La calidad de los equipos y el número de operarios incide significativamente en el tiempo de tratamiento, especialmente cuando nos hallamos ante importantes cantidades de información, por lo que debe ser considerada igualmente.
- **OUTPUT:** La resolución de las salidas digitalizadas, directamente relacionada con el tamaño de celda de los datos raster, debe ser considerada especialmente, pues de ella dependen los tiempos de digitalización y procesado, y puede condicionar negativamente el desarrollo del resto del proyecto.

#### 3.1.2.6. Captura indirecta de datos vectoriales

La obtención indirecta de datos vectoriales supone la aplicación de dos técnicas:

**Digitalización manual:** Es el método más simple y más barato para la vectorización de mapas. Consiste en el uso de tabletas digitalizadoras sobre mapas en papel manejadas por un operario que recorre los elementos cartográficos con un cursor que trasmite la información al ordenador. Si se trata de grandes cantidades de información, este proceso puede resultar laborioso y siempre está sometido a errores por parte del operador. Para cuantificarlo en términos de esfuerzo se deben considerar las capas de información a digitalizar, la escala y la superficie del mapa, además del número de operarios disponible.

**Digitalización en ordenador:** A diferencia del método anterior consiste en el paso de imágenes raster a vectorial por medio del procesado automático o manual en el ordenador. Así, se parte de imágenes ya digitalizadas de la zona de estudio, que son la fuente de datos vectoriales para el SIG. La automatización del proceso agiliza notablemente el producto final, pero comete numerosos errores y debe ser supervisada pormenorizadamente. Cuando el proceso lo lleva a cabo el operador se realiza de manera mucho más lenta, pero resulta más fiable. Los indicadores de esfuerzo propuestos son la resolución de la imagen raster, equipos, extensión y escala, así como las capas a vectorizar y el número de puestos (ordenadores) destinados a tal fin.

### 3.1.3. Software

En este apartado debe considerarse el tipo de software empleado en el SIG, entendiendo como tal el encargado de la gestión integral de la información, concepto que trasciende al de los SGBD anteriormente citados, ya que implica la capacidad de tratamiento de la topología y supone el entorno de relación del usuario con el sistema.

Figura 5: Algunas plataformas de desarrollo de software SIG



En el mercado existen numerosas soluciones de software para Sistemas de Información Geográfica, que pueden ser útiles para proyectos de poca envergadura o escasa especialización, pero si se trata de desarrollos de un cierto calado, es necesario recurrir al desarrollo del software, bien sobre los propios programas comerciales, que ofrecen esta posibilidad, o bien partiendo desde cero, obteniendo un producto realmente adecuado a las necesidades del proyecto, pero en ocasiones más costoso. Si nos referimos a este segundo caso deben considerarse las técnicas de estimación de esfuerzo en desarrollo de software, de las que existe abundante literatura y que no son objeto del presente estudio.

### 3.2. Fase de estimación

La estimación del tamaño y del esfuerzo resultan especialmente importantes en la planificación de un proyecto. Para apoyar esta difícil tarea, se han desarrollado varios métodos que han encontrado aceptación comercial en forma creciente en la planificación del desarrollo de software.

A continuación se citan las principales técnicas de estimación de software; en referencia a su posible aplicación a nuestro ámbito de estudio resulta necesario destacar los siguientes aspectos (Gracia, 2010):

- **Puntos de función:** Se trata de un método que data de 1979. Sus objetivos son medir lo que el usuario pide independientemente de la tecnología proporcionando un determinado factor de normalización para la comparación de distintos software. Se trata de un método ideado para cuantificar el tamaño y la complejidad de un sistema software en términos de las funciones de usuario que este desarrolla (o desarrollará). Esto hace que la medida sea independiente del lenguaje o herramienta utilizada en el desarrollo del proyecto. En su origen, el análisis por puntos de función está diseñado para medir aplicaciones de negocios por lo que no resulta apropiado para otro tipo de aplicaciones como aplicaciones técnicas o científicas. Los algoritmos complejos que se aplican en la gestión de datos espaciales suponen un inconveniente para el método de puntos de función.
- **Puntos de casos de uso:** es un método de estimación de esfuerzo a partir de sus casos de uso. Fue desarrollado por Gustav Kerner en 1993, basándose en el método de puntos de función. Ha sido analizado posteriormente en otros estudios. Este método utiliza los actores y casos de uso relevados para estimar el esfuerzo del desarrollo a estudiar. A los casos de uso se les asigna una jerarquía compleja basada en las transacciones, de interacción entre el usuario y el sistema, mientras que a los actores se les asigna una categorización basada en su tipo, es decir, si son interfaces con usuarios u otros

sistemas. También se utilizan factores de entorno y de complejidad técnica para ajustar el resultado. Dada la carencia de referencias de usos de este método en proyectos SIG, surgen problemas de definición de los casos de estudio que dificultan *a priori* su aplicación de forma sistemática.

- **COCOMO**: es un modelo matemático de base empírica. Incluye tres submodelos, cada uno ofrece un nivel de detalle y aproximación, cada vez mayor, a medida que avanza el proceso de desarrollo del software: básico, intermedio y detallado. Este modelo fue desarrollado por Barry W. Boehm a finales de los años 70, exponiéndolo detalladamente en su libro "Software Engineering Economics" (Prentice-Hall, 1981). Este modelo depende ampliamente de la adaptación del mismo a las necesidades de la organización, a través de datos históricos; los cuales no son fácilmente comparables por analogía en el ámbito de los Sistemas de Información Geográfica, dada la heterogeneidad de este tipo de proyectos.

Se trata, pues, de técnicas típicas de estimación de esfuerzos en proyectos de software, que en el caso de los SIG deben ser adaptadas para considerar las componentes e indicadores definidos en la fase anterior.

### 3.3. Fase de integración

Se define esta fase como aquella en la que se sintetizan y adecúan a la totalidad del Sistema de información Geográfica las técnicas de estimación empleadas anteriormente sobre la base de componentes.

De este modo, teniendo en cuenta los resultados de la estimación de esfuerzo por componentes realizada en la fase anterior se obtendrá un valor global de esfuerzo para cada proyecto, así como una estimación del coste del mismo.

Como resultado del proceso se definirán unos parámetros de ajuste cuya efectividad debe ser testada en varios casos de estudio.

### 3.4. Validación del modelo

La validación de la metodología propuesta requiere para su realización la disponibilidad de información en detalle de un conjunto de proyectos SIG que permita comparar los resultados obtenidos en la estimación con los reales alcanzados en la ejecución posterior. Sería conveniente por tanto disponer de un conjunto de datos de proyectos SIG lo más numeroso posible para garantizar la diversidad de características y la consistencia de la metodología. Sin embargo, no se ha localizado ninguna base de datos de proyectos SIG que disponga de la información necesaria. Para solventar esta dificultad se inició la recopilación y categorización de una serie de casos de estudio a partir de información procedente de distintas empresas especializadas en el desarrollo de sistemas SIG. De esta forma se espera obtener un conjunto de datos que permita probar y validar el modelo propuesto en un amplio rango fuentes y tecnologías de diversa complejidad.

## 4. Conclusiones

Las principales conclusiones de este trabajo son las siguientes:

- La estimación de los costes de desarrollo de los proyectos de Sistemas de Información Geográfica presentan numerosas dificultades debidas, entre otras características, a la heterogeneidad de los datos, a una significativa variedad de tamaños, a la multitud de tecnologías disponibles y la generalidad de aplicaciones y necesidades que deben cubrir

este tipo de proyectos. Todo ello ha hecho que no se disponga de métodos formalizados para la planificación de estos sistemas.

- Para abordar este problema se propone una metodología basada en la identificación inicial de los componentes, su caracterización y posterior estimación, para finalmente integrar los resultados generando un cómputo del esfuerzo global.
- La evaluación de los resultados del modelo propuesto requiere la aplicación sobre proyectos SIG catalogados y jerarquizados. Para ello se ha iniciado una labor de recopilación de sistemas-tipo, generando una base de datos de proyectos SIG que permita evaluar y validar la metodología de estimación de esfuerzos.

### **Bibliografía y Referencias.**

Allan J. Albrecht and John E. Gaffney, "Software Function, Lines of Code, and Development Effort Prediction: A Software Science Validation", IEEE Transactions on Software Engineering, vol SE-9, No 6, November 1983.

Alonso Sarría, Francisco "Introducción a los Sistemas de Información Geográfica". 2001.

Bestebreurtje, J.G.A. GIS Project Management. Manchester Metropolitan University. 1997.

Calvo Melero, Miguel. "Sistemas de información geográfica digitales". IVAP, 1.993.

Charles R. Symons, "Function point Analysis: Difficulties and Improvements", IEEE Transactions on Software Engineering, vol 14, no1, January 1988.

David J Maguire, Michael F Goodchild y David Rhind "Geographical Information Systems, Principles and applications". Longman Scientific & Technical. 1994.

Falkner E., Morgan D., "Aerial Mapping: Methods and Applications" 2nd Ed.CRC Press Company. 2002

Felícísimo, Ángel Manuel "Modelos Digitales del Terreno. Introducción y Aplicaciones en las Ciencias Ambientales". Pentalfa Ediciones. Oviedo, 1994.

Gracia Luis M. Blog personal. <http://unpocodejava.wordpress.com/> 2010.

Harmon J. E., Anderson S. J. "The Design and Implementation of Geographic Information Systems". Wiley, 2003

Marcela P., Varas C. Modelo de Gestión de Proyectos Software: Estimación del Esfuerzo de Desarrollo. Marcela P., Varas C. 1995

Michael Zeiler. "Modeling Our World". Environmental Systems Research Institute, 1999.

Moles "Tecnología de los sistemas de información geográfica". Ed/ra-ma, 1995.

Paul A. Longley, Michael F. Goodchild, David J. Maguire, David W. Rhind. "Geographical Information Systems and Science". Wiley. 2005.

### **Correspondencia** (Para más información contacte con):

Área de Proyectos de Ingeniería. Universidad de Oviedo

Phone: +34 985 10 42 72

Fax: + 34 985 10 42 56

E-mail : [mesa@api.uniovi.es](mailto:mesa@api.uniovi.es)

URL : [www.api.uniovi.es](http://www.api.uniovi.es)