



Herramientas para la Modernización de Catastro y Registro

Módulo 6: Aspecto Tecnológico II

1. El Funcionamiento de un GPS

INTRODUCCIÓN

La Geodesia Espacial se basa en la observación o recepción de señales formadas por radiaciones del espectro electromagnético, procedentes de objetos que no estén físicamente ligados al suelo terrestre.

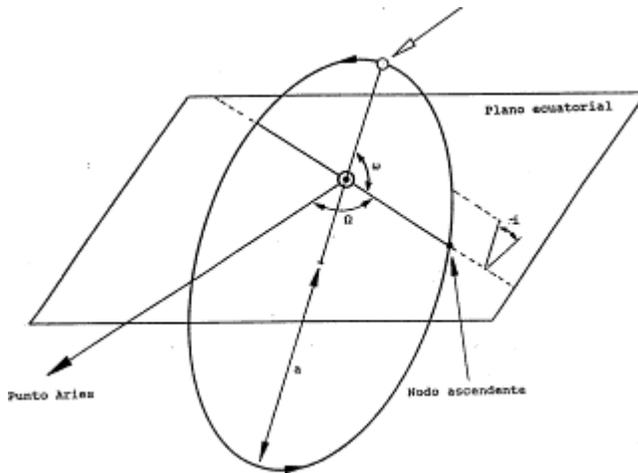
La Geodesia Espacial utiliza fundamentalmente satélites artificiales, sus principales protagonistas.

El sistema GPS (*Global Positioning System* - Sistema de Posicionamiento Global) ha revolucionado el mundo de la topografía y la geodesia.

Antes de empezar con la descripción de los diferentes sistemas, es conveniente hacer ciertas puntualizaciones como las concernientes a parámetros orbitales, tipos de satélites y sistemas de referencia, así como el planteamiento general del posicionamiento por satélite y sus aplicaciones.

En textos técnicos en inglés un satélite se denomina por la sigla **SV** (**Spacial Vehicle** – **Vehículo Espacial**).

ÓRBITAS



Siempre se necesitan seis parámetros para definir una órbita ideal y el satélite en ella. Si bien pueden emplearse otros, los más comúnmente usados son los Keplerianos. Con dos parámetros se fija en el espacio el plano orbital respecto al ecuatorial: **orientación (Ω) e inclinación (i).**

Sobre el plano así establecido se sitúa la elipse de la órbita:

Con tres parámetros más se determina su **orientación (ω)**, **tamaño (a)** y **excentricidad (e)**: Ya está definida la órbita elíptica en el espacio. La velocidad instantánea del satélite y su período son determinables por las leyes de Kepler; sólo resta establecer un último parámetro que fije el satélite en un punto concreto de la órbita en un momento dado, por ejemplo, en el momento del **paso del satélite por el perigeo (t)**.

A este planteamiento ideal hay que añadir mucha más información para poder predecir la posición instantánea real del satélite, que es diferente de la teórica calculada con los 6 parámetros básicos, por una serie de perturbaciones orbitales que alteran su trayectoria. Estas son: anomalías gravitacionales originadas por el efecto de terceros cuerpos, como la luna y el sol, irregularidades y variaciones del campo gravitatorio terrestre, fricción atmosférica y presión de la radiación.



En consecuencia, si deseamos establecer la posición de un satélite con una gran precisión, las efemérides necesarias no puedan limitarse a las keplerianas (o alternativamente a sólo 6 básicas). En el caso del GPS, las 6 keplerianas básicas descritas se complementan con otros 11 parámetros más:

Símbolo	Unidad	Parámetro
AODE	segundos	Antigüedad de la información de efemérides
C_n	metros	Amplitud de la corrección armónica senoidal del radio
Afl	n radianes / s	Diferencia del movimiento medio
M0	n radianes	Anomalía media en el momento de referencia
c_{uc}	radianes	Amplitud de la corrección armónica cosenoidal del argumento de la latitud
e	Adimensional	Excentricidad (0,03)
c_{us}	radianes	Amplitud de la corrección armónica senoidal del argumento de la latitud
(A)1/2	metros 1/2	Raíz cuadrada del semieje mayor
t_{Oe}	segundos	Tiempo de referencia de efemérides - valor máximo 604
C_{ic}	radianes	Amplitud de la corrección armónica cosenoidal del ángulo de inclinación
120	a radianes	Ascensión recta en el momento de referencia
C_{js}	n radianes	Amplitud de la corrección armónica senoidal del ángulo de inclinación
$\dot{\Omega}_0$	n radianes	Ángulo de inclinación en el momento de referencia
C_{rc}	metros	Amplitud de la corrección armónica cosenoidal del radio orbital
W	n radianes	Argumento del perigeo
OMEGADOT	n radianes / s	Razón del cambio en la ascensión recta
IDOT	n radianes / s	Razón del cambio en el ángulo de inclinación



EFEMÉRIDES RADIODIFUNDIDAS

SISTEMAS DE REFERENCIA

Es evidente que la utilización de un satélite, especialmente en las aplicaciones que más nos son afines, exige un conocimiento preciso de su situación espacio-temporal en forma de coordenadas concretas, lo que se resuelve con el conocimiento de las efemérides, siempre que se disponga del sistema de referencia adecuado.

Para los satélites se usa un sistema inercial cartesiano geocéntrico, en el que un eje es el terrestre, otro es la dirección al punto Vernal y el tercero es ortogonal a ambos, llamado en textos en inglés ***Conventional Celestial Reference System (CCRS)***.

La geometría de una órbita referida a un sistema de este tipo es constante, y por ello es fácil referir la posición de un satélite como X, Y y Z, pero el problema es que las coordenadas del punto de la Tierra en el que se realiza una observación variarían continuamente.

Es fácil pasar al equivalente no inercial, en el que el eje que se dirigía al punto Vernal se hace pasar por el meridiano de Greenwich, por ejemplo; sólo hay que tener en cuenta la rotación terrestre.

En este nuevo sistema cartesiano geocéntrico, las coordenadas del terreno son constantes, y es llamado en inglés ***Earth Centered Earth Fixed (ECEF)*** o ***Conventional Terrestrial Reference System (CTRS)***. Se pueden transformar fácilmente las anteriores



coordenadas de las posiciones de los satélites a este sistema, y además se pueden obtener cómodamente las de puntos terrestres situados gracias a ellos.

Desde este sistema de referencia podemos pasar a otros, como a un elipsoide, o a UTM, por ejemplo, como veremos más adelante.

POSICIONAMIENTO

En general, cuando desde un punto se miden tres direcciones o distancias no coplanarias a satélites de posición conocida, se fija la situación del punto respecto a la de los satélites. En este caso sólo hay una solución: con más direcciones o distancias hay soluciones múltiples (redundancia observacional), lo que permite detectar observaciones erróneas, localizarlas y finalmente ajustar la solución final, a medida que aumenta la redundancia.

El problema siempre puede invertirse: en sucesivas observaciones a los satélites desde puntos conocidos, se pueden establecer los parámetros orbitales, lo que se hace desde las estaciones de control.

Para expresarlo de una forma elemental, si sabemos dónde están los satélites, podemos situarnos; si sabemos nuestra posición, podemos situar los satélites (o determinar sus efemérides).

Como en cada proceso se pierde precisión, si la posición de una estación de control no está establecida con un rango superior, no se podrán situar los satélites (determinar sus efemérides) con la precisión necesaria para después utilizar estas efemérides; por esto los puntos de seguimiento o control siempre se establecen previamente con un sistema SLR, o aún mejor, con un sistema VLBI, descritos más adelante.



Es esencial mencionar que para poder prever las posiciones conocidas de los satélites es preciso tener establecida una cuarta incógnita: el tiempo, en función del cual las efemérides nos permiten calcular las posiciones de los satélites.

PROPAGACIÓN DE EMISIONES RADIOELÉCTRICAS

Una onda electromagnética que provenga del espacio debe atravesar tres zonas características antes de alcanzar un receptor estacionado sobre la superficie terrestre: el vacío, la ionosfera y la troposfera.

El retardo es el incremento que sufre el tiempo de propagación de una señal electromagnética entre dos puntos al efectuarse el tránsito por un medio que no sea el vacío, en vez de hacerlo por el vacío. Se debe a dos factores: la velocidad de propagación es menor y la trayectoria (si no pasa por el cenit) aumenta su longitud al curvarse por refracción y ser envolvente de la recta que une los puntos origen y destino de la señal.

En el vacío, el retardo es inexistente.

En la ionosfera, que está entre 100 y 1,000 km de altitud, el retardo es proporcional al número total de electrones libres encontrados por la señal en su camino. Varía para cada punto concreto de recepción según su latitud y la dirección y momento de la observación. Pueden usarse modelos ionosféricos para establecer el retardo, pero con cierta indeterminación, dado el comportamiento más bien irregular e imprevisible de la ionosfera.



Para resolver el problema, en lo que atañe a la recepción de señales provenientes de un satélite, se puede emplear el artificio de utilizar dos frecuencias diferentes y razonablemente separadas dentro de la banda de trabajo.

Observando el retardo diferencial entre ambas, tanto mayor cuanto mayor sea el retardo ionosférico sufrido, se puede deducir indirectamente el retardo total, con una precisión aceptable.

En la troposfera el retardo se puede conocer utilizando modelos matemáticos, en función de tomas locales de parámetros meteorológicos: temperatura, humedad y presión. En distancias cortas y medias, en las que el retardo ionosférico no es significativo, esto puede obviarse dado que la diferencia de condiciones locales entre ambos puntos genera correcciones irrelevantes.

SISTEMAS ACTUALES

Los sistemas actualmente operativos en nuestro ámbito científico son: SLR, VLBI, TRANSIT y GPS.

- El **SLR (Satellite Laser Ranging)** es un sistema de medida directa de distancias por pulso láser a satélites provistos de prismas de reflexión total.
- La **VLBI (Very Long Baseline Interferometry)** es una técnica que permite calcular con precisión centimétrica la distancia entre los centros radioeléctricos de dos o más radiotelescopios. Se observan las señales provenientes de cuásares extra-galácticos en períodos simultáneos, comparando interferométricamente las señales recibidas.



Estas dos técnicas permiten el establecimiento de una red mundial increíblemente precisa, que actualmente constituye un marco de referencia absoluto para todas las ciencias que lo necesiten.

El sistema **TRANSIT**, genéricamente conocido como **DOPPLER**, se basa en la medición de la variación de distancias a satélites mediante el corrimiento Doppler de la frecuencia de las señales recibidas.

El GPS, al que se dedican los siguientes capítulos, puede trabajar en sistema Doppler, en medida directa de distancias o en comparación de fase de frecuencia. Es un sistema flexible y poderoso que ya ha cambiado multitud de técnicas de observación topográfica y geodésica, como lo hizo la distanciometría electrónica en su tiempo, si bien en este caso el fenómeno ha ocurrido con una rapidez muy superior.

DESCRIPCIÓN DEL SISTEMA

En 1973 se empezó a desarrollar el proyecto de la constelación **NAVSTAR** (*NAVigation Satellite Timing And Ranging* - Satélite de Navegación, Cronometría y Distanciometría) por el Departamento de Defensa (DoD) de los Estados Unidos para poder hacer navegación más precisa en tierra, mar o aire, de forma continua y en tiempo real, en toda condición meteorológica y en un sistema unificado de cobertura global, que se llamó GPS; el 22 de febrero de 1978 se lanzó el primer satélite.

El sistema es fundamentalmente militar, pero los estamentos científicos civiles empezaron a intentar utilizarlo desde el comienzo de la década de los 80, porque el GPS es algo más que un mero sistema de navegación; con él se pueden hacer diferentes tipos de



posicionamiento, del máximo interés en nuestras técnicas por la elevada precisión alcanzable.

Los posicionamientos realizables con el GPS se pueden dividir en dos, dependiendo de si el receptor se mueve o no.

- Si el receptor se mueve, el posicionamiento se llama **dinámico** y es usado, por ejemplo, en navegación.
- Si el receptor no se mueve, el posicionamiento se llama **estático** y es usado, por ejemplo, en topografía.

También se puede hacer una clasificación según lo que se quiera referir a la o las posiciones halladas:

- Si las coordenadas de un punto singular se determinan respecto a un sistema de coordenadas previamente definido, el posicionamiento se llama **absoluto**.
- Si las coordenadas se determinan respecto a otro punto, que constituiría la referencia de un sistema local de coordenadas, el posicionamiento se llama **diferencial o relativo**.

La distancimetría geodésica de Primer Orden ha sido definitivamente desplazada. Actualmente han desaparecido los problemas de costo y operatividad que en los últimos 10 años habían contenido su plena difusión en otras áreas, según se iba completando la constelación, ya ultimada. En geodesia y topografía es absolutamente operativo, ofreciendo en muchas aplicaciones la solución más rentable posible.

Esta capacidad de situar puntos se realiza en el marco de un sistema de referencia global extraordinariamente homogéneo, garantizando, además de observaciones breves y fáciles, una coherencia y uniformidad de escala y orientación sin precedentes hasta ahora.



Básicamente hay dos tipos de receptores a considerar: los expeditos, pequeños, ligeros y baratos, que observan normalmente en posicionamiento absoluto por pseudo distancias, y los topográfico-geodésicos, mayores, que exigen estacionamiento más preciso, observando normalmente en posicionamiento relativo por medida de fase. Las aplicaciones de estos últimos, que son los dirigidos al mundo de la topografía y la geodesia, podrían sintetizarse en el siguiente breve resumen:

- Establecimiento de redes de control con precisión mejor que 4 ppm, a partir de las efemérides radiodifundidas por los propios satélites, o mejor que 1 ppm, a partir de efemérides precisas.
- Determinación precisa de perfiles del geoide sobre líneas de nivelación de alta precisión.
- Densificación de redes geodésicas con precisiones mejores que ± 1 cm en las tres coordenadas, sobre lados de 2 a 5 km.
- Apoyo fotogramétrico en escalas medias y altas con excelente rendimiento en cualquier tipo de terreno, y precisión mejor que ± 20 cm con observación GPS corta.
- Establecimiento de bases de replanteo de alta precisión en obras lineales de largo recorrido, como carreteras, ferrocarriles, canales, etc.
- Establecimiento de redes de alta precisión para el replanteo de grandes obras de ingeniería como túneles, puentes, presas, etc.
- Determinación expedita de redes eléctricas, telefónicas, de conducción de agua, oleoductos, etc.
- Localización de obras hidráulicas en cuencas hidrográficas.
- Estudio de la evolución fluvial
- Batimetrías con autocontrol de navegación, sin tener que considerar correcciones de oleaje y marea, con gran precisión y automatización.



- Gran flexibilidad y fiabilidad en la adquisición de la información necesaria para actualizar cualquier GIS, permitiendo su inmediato tratamiento digital y/o gráfico.
- Inmediata localización y delimitación de zonas afectadas en grandes desastres, como incendios, inundaciones, contaminaciones, epidemias, plagas, etc.
- Localización y control de vehículos de auxilio y servicio.
- Inventario expedito de redes viales de cualquier tipo, como fluviales, de carreteras, caminos, pistas forestales, etc.
- Estudio de mareas en puntos alejados de tierra.

CONSTITUCIÓN

El sistema GPS puede considerarse dividido en tres sectores fundamentales:

El espacial, el de control y el de usuarios.

SECTOR ESPACIAL



Está constituido por la constelación de satélites NAVSTAR (figura n° 3), que consta de 6 órbitas, prácticamente circulares, con inclinación de 55° y uniformemente distribuidas en el plano del ecuador, con 4 satélites por órbita, uniformemente distribuidos y con altitud de 20,180 km, lo que corresponde a un período de 12 horas de tiempo sidéreo. Es decir, el satélite completa dos órbitas exactas cada 360° de giro de la Tierra, por lo que la trayectoria terrestre del satélite (traza del vector geocéntrico del satélite sobre la superficie terrestre) se repite exactamente cada día sidéreo.



Los 24 satélites y sus 4 de reserva de la constelación NAVSTAR circundan la tierra distribuidos de tal manera que en cada punto de la superficie terrestre se tiene la posibilidad de leer la señal de al menos cuatro satélites. Esto es muy importante, porque se necesitan al menos cuatro satélites para conocer la posición del observador, y que estos se dispongan con un ángulo de elevación sobre el horizonte superior a 15°; no obstante, casi siempre son más de cuatro los satélites 'visibles'.

Los satélites envían señales en la región de radio del espectro electromagnético. La señal en sí es muy compleja. Está formada por varios componentes que se estructuran sobre una señal principal con frecuencia de 10'23 MHz. A partir de esta señal principal, y derivada de ella, se producen los dos componentes principales de la señal: las portadoras (*carriers*).

Esto significa que la configuración local de la constelación se repite 3 m 56 s antes cada día solar, lo que representa casi media hora semanal o unas dos horas mensuales de adelanto.

En cualquier punto y momento hay entre 6 y 11 satélites observables, con geometría favorable.

Al ser un sistema nacido de la investigación militar y con una importancia geoestratégica obvia, el gobierno de los Estados Unidos se preocupó mucho de que se pudiera garantizar el uso adecuado. En principio, se degradaba la señal intencionalmente para que los receptores civiles tuvieran un error mínimo que hiciera inapropiado su uso para aplicaciones militares. Era lo que se llamaba la Disponibilidad Selectiva (*Selective Availability*) que condicionaba las lecturas a un error mínimo de 100 m. mediante la modificación de los datos de tiempo del satélite (reportados por los relojes atómicos a bordo) y alterando las efemérides de los satélites.



El 1° de mayo de 2001, la administración Clinton decidió eliminar esta fuente de error intencionada, dada la importancia económica que estaba tomando el GPS; a partir de ese momento, la precisión del sistema se mejoró notablemente, tal y como se puede ver en los siguientes gráficos, en los que se documenta el momento en que se eliminó la Disponibilidad Selectiva (S/A):

No obstante, existe otro modo de anular la señal en caso de existir un conflicto bélico en alguna región del planeta. A través del procedimiento de *Anti-Spoofing* (A-S), los gestores del sistema pueden encriptar totalmente la señal. Mediante el uso de un código adicional de alto secreto (denominado W), se consigue encriptar el código P, que pasa a denominarse entonces código Y; este código Y sólo se puede leer con receptores GPS militares autorizados, con lo que se garantiza la exclusividad mediante una denegación selectiva del servicio en zonas de conflicto.

SATÉLITES

Los satélites NAVSTAR pesan menos de una tonelada en el momento de la inserción en órbita de servicio. Todos incorporan osciladores atómicos.

Tienen una serie de antenas emisoras que funcionan en la banda L del espectro, y son las encargadas de enviar a la superficie terrestre las señales que se reciben.

En las técnicas GPS, los satélites se identifican por su ruido pseudo aleatorio (**Pseudo Random Noise - PRN**), característico y exclusivo de cada satélite NAVSTAR en particular, que será descrito más adelante.

Para la fuente de frecuencia a bordo de un satélite, la terminología GPS adopta el término reloj (*clock*) en lugar del más habitual de "oscilador".



El funcionamiento de un reloj u oscilador atómico se basa en la transición entre niveles de energía de átomos concretos previamente excitados. La transición produce una oscilación de frecuencia muy precisa y estable.

Los relojes de los satélites son osciladores atómicos, que pueden ser de rubidio, con precisiones de 10^{-12} o de cesio, de 10^{-13}

Debe reflexionarse qué representa una precisión de 10^{-13} como la mencionada:

10^{-13} segundos es lo que la luz invierte en recorrer 30 milésimas de milímetro, y esa precisión representa una variación de un segundo en 300 000 años.

El reloj u oscilador de servicio de un satélite GPS provee una frecuencia fundamental de 10,23 MHz, sobre la que se estructura todo el conjunto de la señal radiodifundida por el satélite.

TIEMPO

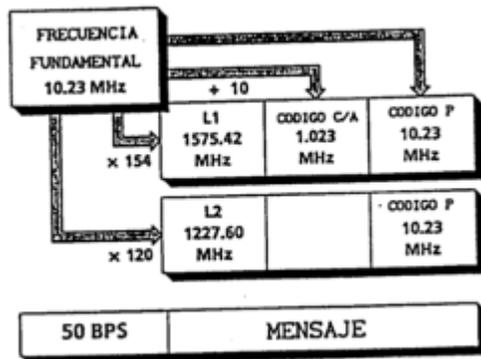
La escala de tiempo usada en el sistema GPS no coincide exactamente con ninguna de las actualmente definidas y utilizadas.

El US Naval Observatory dispone de 25 relojes atómicos con los que establece su propia escala interna de tiempo atómico UTC (Tiempo Universal Coordinado), identificada como **UTC (USNO)**.

En técnica de satélites es lógico usar una escala uniforme, independiente de las variaciones rotacionales terrestres, dado que el movimiento de aquellos no depende de éstas. Para ello, el US Naval Observatory también establece otra escala de tiempo atómico, que llama **GPS Time**, cuya unidad es el segundo atómico internacional.

El origen de la escala GPS se ha fijado como coincidente con el UTC a las 0 h del 6 de enero de 1980. Como en ese momento el TAI (Tiempo Atómico Internacional) difería del UTC en 19 segundos, el GPS Time es equivalente al TAI menos 19 segundos, y así ha de mantenerse dado que ambas escalas son atómicas y uniformes, y por tanto paralelas.

PORTADORAS



El satélite emite sobre dos ondas portadoras (figura n° 5): una es el resultado de multiplicar la fundamental (10,23 MHz) por 154: **1575,42 MHz, que es llamada L1 (1 de 19,05 cm)**. La otra usa un factor de 120: **1227,60 MHz, y se llama L2 (X de 24,45 cm)**. La L es porque los valores usados están en la banda L de radiofrecuencias,

que abarca desde 1 GHz a 2 GHz

El hecho de disponer de 2 frecuencias permite determinar, en caso necesario, el retardo ionosférico por comparación de sus diferentes retardos, técnica descrita previamente.

La banda L del espectro es la que presenta mejor transparencia atmosférica, lo cual es muy importante para la precisión del sistema.

CÓDIGOS

Sobre las portadoras L1 y L2 antes descritas, se envían dos códigos y un mensaje por modulación, cuya base también es la frecuencia fundamental 10,23 MHz

1. El primer código, llamado **C/A** (*Course/Acquisition*) o **S** (*Standard*), es una moduladora usando la frecuencia fundamental dividida por 10, o sea 1,023 MHz.
2. El segundo código, llamado **P** (*Precise*), modula directamente con la fundamental



de 10,23 MHz.

3. Finalmente, el **mensaje** se envía modulando en la baja frecuencia de 50 Hz (es el factor $204\ 600^{-1}$ de la fundamental).

Los códigos sirven fundamentalmente para posicionamiento absoluto y son usados principalmente en navegación. El C/A ofrece precisiones nominales decamétricas y se usa en el servicio de posicionamiento estándar SPS (*Standard Positioning Service*). El código P, actualmente sustituido por su equivalente secreto Y (lo que se llama "**Anti-Spoofing**" - AS) ofrece precisiones nominales métricas y se usa en el servicio de posicionamiento preciso PPS (*Precise Positioning Service*). El mensaje aporta toda la información necesaria para los usuarios del sistema.

Los códigos y el mensaje consisten en una secuencia de determinada longitud formada por dígitos binarios (ceros y unos) que se denominan chips, generados según desarrollos polinómicos conocidos.

Los códigos son generados mediante un sistema llamado **Tapped Feedback Shift Registers** (TFSR).

La modulación de las portadoras con estos códigos (mediante modulación binaria bifase) genera un ruido que aparentemente no sigue ninguna ley y parece aleatorio, pero como en realidad la secuencia está establecida mediante un desarrollo polinómico, el ruido se llama pseudo aleatorio (*Pseudo Random Noise* - **PRN**) y puede correlacionarse con una réplica generada por el receptor en tierra.

Cada código tiene una configuración propia para cada satélite en particular, y constituye el PRN característico con que éste es identificado.



Sobre ambas portadoras se modula el código P y el mensaje; sobre la L1, se modula además el código C/A.

MENSAJE

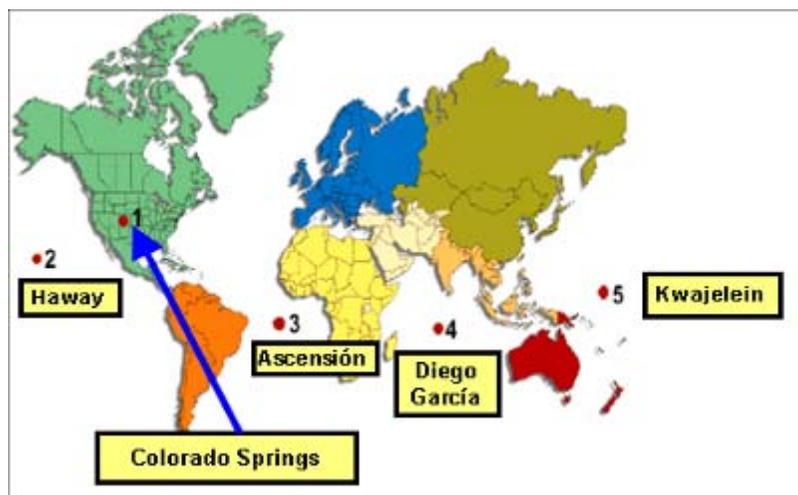
La emisión del mensaje tiene una duración de 12 m 30 s, repetida continuamente. El **mensaje** contiene las efemérides radiodifundidas que se usan para la obtención de resultados, un modelo ionosférico para usuarios mono-frecuencia, el almanaque (que es una información expedita de las órbitas de todos los satélites usada para planificar observaciones, el estado de relojes de los satélites, la condición de cada satélite (llamada salud (**Health**) en terminología GPS: puede estar sano (**OK**) o enfermo (**Unhealthy**)), la antigüedad de la información transmitida, y otras indicaciones.

Los propietarios del sistema han querido limitar la precisión obtenible a usuarios no autorizados y han aplicado lo que se llama "**Disponibilidad Selectiva**" (*Selective Availability, SA*). Esta consiste en falsear ligeramente los parámetros orbitales y las efemérides. Su efecto es que el posicionamiento de los satélites sufre errores del orden de 100 m, error que se transmite en el posicionamiento absoluto de un punto en la superficie terrestre.

Afortunadamente, para los usuarios de las Ciencias Geográficas que usan la medida de fase, este error absoluto de 100 m, considerando la figura geodésica satélites-receptores con lados de 25,000 km, repercute en un error relativo del orden de 4 ppm, que es la limitación que se tiene en el posicionamiento relativo. Algo impensable hace tan sólo unas décadas en posicionamiento de puntos no necesariamente inter-visibles y en cualquier condición.

SECTOR DE CONTROL

Hay cinco estaciones oficiales de seguimiento para los satélites de la constelación NAVSTAR: la estación central o maestra de Colorado Springs, y las secundarias de Ascensión -Atlántico Sur-, Diego García -Índico-, Kwajalein -Pacífico occidental- y Hawái -Pacífico oriental-.



Están espaciadas regularmente en longitud y tienen coordenadas muy precisas; recibiendo continuamente las señales de los satélites que estén sobre el horizonte local en las dos frecuencias, obtienen la información necesaria para establecer con gran precisión las órbitas de todos satélites.

Todos los datos se envían a la central, donde se procesan y se calculan las efemérides, los estados de relojes y toda la información que luego se transmite y almacena en la memoria de cada satélite para su radiodifusión.

SECTOR USUARIO

Son los instrumentos utilizados para hallar coordenadas de un punto o hacer navegación, usando las señales radiodifundidas desde los satélites NAVSTAR.



El equipo esencial está formado por antena y receptor.

La antena, considerada como conjunto, consta siempre del elemento físico de la antena receptora, un preamplificador y la bancada de estacionamiento.

El receptor consta de una sección de recepción de radiofrecuencias, con diferentes canales (entre 5 -mínimo operativo- y 12 -máximo observable-) para seguir simultáneamente diferentes satélites, un procesador interno con su correspondiente soporte lógico (firmware y software), que entre otras funciones puede generar réplicas del código C/A y/o del P, una unidad de memoria con soporte sólido o de tarjeta, conectores y salidas; como periféricos, un teclado de control y una pantalla de comunicación. Finalmente una fuente de alimentación, externa o interna, pero siempre con toma exterior para poder conmutar de una unidad de alimentación a otra sin interrumpir el funcionamiento del receptor.

ANTENA

La función de una antena es convertir una corriente eléctrica en una radiación electromagnética, y viceversa. La antena (emisora) de los satélites hace lo primero y la de nuestros instrumentos (receptora) lo segundo.

La corriente eléctrica inducida en nuestra antena por las señales radiadas recibidas posee toda la información modulada sobre ellas. Hay diversos tipos de antenas: dipolo, dipolo curvado, cónico-espiral, helicoidal, plana (microstrip), etc.

El punto que realmente se posiciona es el centro radioeléctrico de la antena, y cada tipo tiene sus características propias.



Las antenas GPS deberían tener idealmente un lóbulo de recepción hemisférico. Es decir, deberían recibir con igual sensibilidad en cualquier dirección, desde el cenit hasta el horizonte, pero no por debajo del horizonte, al menos en el tipo de aplicaciones de las Ciencias Geográficas.

Se trata de evitar la recepción de señales reflejadas en el suelo u objetos cercanos, porque empeoran la precisión de observación. Este es un problema clásico en radiotransmisión y distanciometría electromagnética, conocido como efecto multi-camino (**multipath**).

Para evitar este fenómeno era frecuente que las antenas llevaran un plano de tierra, tanto mayor cuanto más antiguas. Actualmente, las antenas se diseñan con un lóbulo de recepción que no permite la recepción de estas señales reflejadas en el suelo. Desde luego, la única forma de evitar la reflexión en objetos próximos, como un rascacielos, por ejemplo, sería apantallar la antena.

La gran ventaja de las antenas GPS es que sus exigencias de estacionamiento son muy inferiores a las de los instrumentos de observación clásica: la altísima estabilidad, imprescindible para observar con un teodolito (pilar de hormigón y base pesada o placa de estacionamiento en observación geodésica, o trípode perfectamente estacionado en topografía, por ejemplo), ya no es necesaria, porque los pequeños desplazamientos, vibraciones o torsiones en nada afectan las observaciones a satélites.

Otra ventaja es que ya no es indispensable la tradicional, cara y estable “monumentación” de hitos geodésicos de hormigón, esencial para una correcta observación con teodolito,



siendo suficiente en general una señal mínima que materialice el punto al que damos coordenadas.

Clásicamente, la antena era un elemento independiente del receptor, al que se unía por un cable. Posteriormente hubo modelos compactos que llevaban la antena montada en la parte superior del propio receptor y permanentemente conectada al mismo, con el inconveniente de que el estacionamiento de la antena (y el equipo al que va unida) ha de hacerse, como máximo, a la altura de manipulación del observador, restando flexibilidad de observación. Si se estaciona alto, su manejo es incómodo, y si se estaciona bajo, hay riesgo de pérdida de ciclos de algún satélite con poca altura de horizonte, si el observador inadvertidamente se interpone en el camino de recepción; si se observa con climatología adversa, se renuncia a la comodidad que representa instalar el receptor en el interior del vehículo, dejando la antena en el exterior, que siempre es a prueba de lluvia.

Las últimas generaciones de receptores "multiuso" ofrecen una configuración que aúna todas las ventajas: el receptor, la antena, y frecuentemente la fuente de alimentación, se integran en un conjunto reducido, pero sin periféricos de pantalla y teclado, que se sustituyen por un pequeño ordenador de mano, de tipo comercial, tipo calculadora. El registro de observaciones se hace en tarjeta cambiable o en memoria interna. En ambos casos, después se vuelca a un PC para el procesamiento conjunto.

Los costosos y poderosos receptores geodésicos con las máximas características necesarias para medir lados muy largos, de muchas decenas de kilómetros, siguen conservando una configuración clásica.



ETAPA RECEPTORA

En la antena se han de generar tantas señales como satélites se estén recibiendo. Cada señal necesita un canal o dispositivo electrónico que la procese con independencia de las demás, tras ser separadas y aisladas por el receptor.

CORRELACIÓN/CUADRATURA

Sobre la señal, además de la cuenta Doppler, se pueden hacer dos tipos fundamentales de medidas: **seudo distancias** o **medida de fase**. Su aplicación en medida se describirá en el siguiente capítulo.

SEUDO DISTANCIAS

La medida de seudo distancias se hace sobre la portadora, modulada siempre mediante el método de **correlación**.

Para la correlación hace falta un circuito que seguirá a los códigos, identificándolos y correlacionándolos con una réplica exacta generada por el propio receptor y que permitirá medir las distancias (**seudo distancias**) al satélite en función del tiempo transcurrido entre el momento de la emisión y el de la recepción.

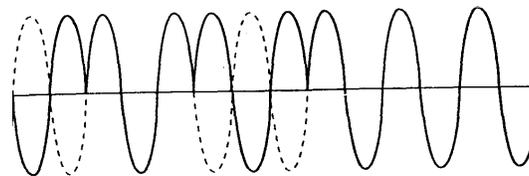
MEDIDA DE FASE

Para la medida de fase se ofrece una alternativa en el proceso que se puede aplicar a la portadora modulada: correlación o cuadratura. El objeto final es disponer de la onda portadora limpia de modulación para poder medir sobre ella, por comparación de fase, cómo evoluciona la distancia satélite-receptor.



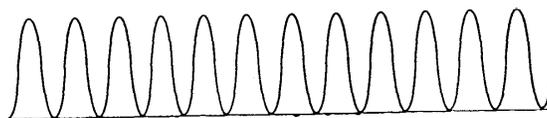
La correlación permite disponer de la portadora original, limpia de modulación, al ser sometida la señal recibida a una "contra-modulación" por el receptor, restaurándose así la portadora original. El proceso de cuadratura es simple: se eleva al cuadrado la amplitud instantánea de la señal recibida.

El satélite modula la portadora multiplicando su amplitud por +1 y por -1; al multiplicar por sí misma la amplitud instantánea de la portadora modulada, la modulación desaparece, ya que $+1^2 = +1$ y $-1^2 = +1$, obteniéndose así un armónico de la portadora, pero totalmente limpio de modulación, con el que pueden hacerse medidas de fase y cuenta de incrementos de ciclos.



$$A * \cos wt$$

Al multiplicarla por sí misma obtendremos



$$(A^2 * \cos(2wt) + 1)/2$$

DEMODULACIÓN POR CUADRATURA

Al no tener información (por haber sido eliminada al eliminar la modulación), es preciso tener acceso a efemérides y estados de reloj de alguna forma, aunque esta información se puede haber adquirido antes de la cuadratura.



Otro problema del método de cuadratura es que como la señal es menor que el ruido, al elevarlas ambas al cuadrado la relación señal/ruido empeora, impidiendo el seguimiento de satélites bajos. La ventaja es que no es preciso conocer, generar y correlacionar los códigos en el receptor.

Este sistema de demodulación se denomina *codeless* (sin código) en inglés, porque se pierden ambos códigos y el mensaje de navegación.

Por correlación o por cuadratura disponemos ya de la portadora libre de modulación. Sobre ella, el circuito comparador de fase adecuado permitirá observar cómo aumenta o disminuye el número de ciclos o longitudes de onda y las fracciones de ciclo o diferencias de fase contenidas en la distancia entre satélite y receptor, lo que permitirá medir incrementos o decrementos en la distancia según varíen la situación relativa entre satélite y receptor.

RELOJ/OSCILADOR

Para hacer todas las labores descritas en la etapa receptora se precisan frecuencias de referencia, obtenidas a partir de la fundamental del reloj oscilador principal del receptor, que normalmente es un oscilador de cuarzo de altísima calidad y estabilidad.

En el futuro, en los receptores GPS de doble frecuencia usados en las más exigentes aplicaciones científicas y oficiales, especialmente para la alta geodesia, probablemente sea estándar el reloj atómico.

En el otro extremo del mercado permanecerán los receptores mono-frecuencia "multiuso", con oscilador a cuarzo, opción barata plena y masivamente usable en aplicaciones topográficas y en las geodésicas de menor orden, locales o de densificación.



INFORMACIÓN

Una vez en funcionamiento, a través de la pantalla y tras las oportunas manipulaciones del teclado, el receptor puede ofrecer al operador una gran cantidad de información sobre el proceso de observación. Aunque varía entre los diferentes modelos, en general se suele disponer de bastantes de las siguientes posibilidades:

- Resultados de la comprobación interna inicial (*start check*).
- Satélites seleccionados manualmente para trabajar con ellos.
- Satélites localizados.
- Satélites en seguimiento.
- Acimut de cada satélite en seguimiento.
- Elevación de cada satélite en seguimiento.
- Número de eventos (*epoch*) seguidos registrados de cada satélite en seguimiento.
- Intensidad de cada señal recibida (relación señal/ruido).
- Condición -salud (*healthy*)- de cada satélite en seguimiento, con información sobre problemas en el mensaje, problemas en portadoras, problemas en moduladoras, falla generalizada, etc., todo ello bajo un código establecido.
- Posición actual (longitud, latitud y altitud).
- Dirección y velocidad del movimiento (en navegación).
- Si se ha dado información del punto de destino: distancia al punto de arribo, tiempo de llegada, desviaciones del curso previsto, etc. (en navegación).
- Diferentes puntos de destino prefijados en la memoria y sus datos (en navegación).
- Bondad de la geometría de observación.
- Progreso de la observación: satélites que se pierden y captan y número de observaciones realizadas a cada uno.
- Opción de observación elegida: estática, cinemática o dinámica, y absoluta o



diferencial; número mínimo de satélites necesarios para registrar, altura de máscara, intervalo entre registros, parámetros fijados tales como altura conocida, etc.

- Nombre y número de sesión que hemos dado a la estación, identificación del operador y notas varias.
- Ocupación de memoria por las observaciones realizadas.
- Estado de la fuente de alimentación, y toda la información útil que el usuario pueda necesitar.

SISTEMAS DE MEDIDA DE DISTANCIA

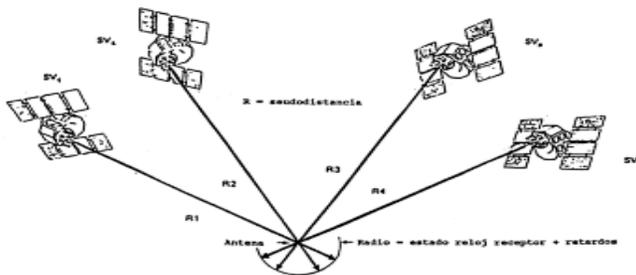
El GPS, como se sabe, es un sistema que permite hacer posicionamiento por medición de distancias entre las antenas emisoras de varios satélites y la receptora de los equipos de observación.

Hay dos posibilidades para medir la distancia entre satélite y usuario: por pseudo distancias o por medida de fase, también llamada diferencia de fase y a veces interferometría, con cierta impropiedad. No mencionaremos el sistema Doppler, también posible, por no utilizarse directamente ni en topografía ni en geodesia

SEUDO DISTANCIAS

El método de **seudo distancias** es exclusivo de la técnica GPS. Es una auténtica trilateralización (o multilateralización) tridimensional que sitúa la estación en la intersección de unas esferas con centro en el satélite y radio de la distancia correspondiente. Este sistema es el usado en navegación y permite el posicionamiento continuo.

El satélite modula sus portadoras con dos códigos: el C/A y el P. El C/A es repetido cada milisegundo.



El receptor tiene en su memoria la estructura del código en cuestión, y genera una réplica exacta. Para sincronizar la réplica con el original recibido, el

instrumento empieza a aplicar un retardo. Cuando la anulación sucede, el tiempo de retardo nos permite calcular una distancia, que no será precisamente la existente ya que, aunque sabemos el momento de la emisión en el satélite, no conocemos el estado del reloj del receptor. Por eso el valor hallado no es una distancia, sino una pseudo distancia.

Para explicar de una forma muy conceptual cómo se resuelve el estado del reloj propio, pensemos que la antena está situada en el centro de una pequeña esfera tangente a todas las esferas sucesivas con centro en cada posición instantánea de satélites y con radio de la correspondiente pseudo distancia observada. La mencionada pequeña esfera con centro en la antena tiene como radio el algebraicamente correspondiente en tiempo/luz al estado A_t del reloj propio.

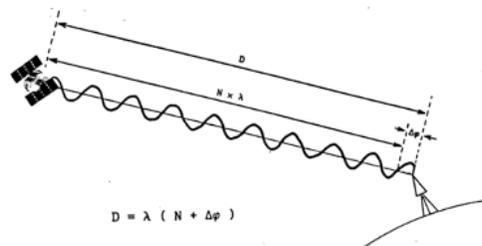
Es evidente que tres esferas dan un solo punto de intersección, siendo por tanto preciso disponer de un cuarto satélite para determinar nuestro estado del reloj. Hay 4 incógnitas: X, Y, Z y A_t .

Para usos topográfico-geodésicos, este método, esencial en navegación por dar resultados en tiempo real, carece de interés dada la precisión obtenible, decamétrica o peor, pero sí se emplea para un breve posicionamiento aproximado, automático en los instrumentos modernos, previo a la observación por medida de fase.

MEDIDA DE FASE

El método de **medida de fase** es el que permite la mayor precisión: una frecuencia de referencia, obtenida del oscilador que controla el receptor, se compara con la portadora desmodulada que se ha conseguido tras la correlación, o bien sobre su armónico conseguido por el método de cuadratura. La base del método es que se controla en fase una emisión radioeléctrica hecha desde el satélite, con frecuencia conocida y desde posición siempre conocida (aunque continuamente variable).

Cuando la onda portadora llega a la antena, habrá recorrido una distancia D , correspondiente a un cierto número entero N de sus longitudes de onda, más una cierta parte de longitud de onda $\Delta\lambda$ (figura n° 9). Así:



MEDIDA DE FASE

El observable es $\Delta\Phi$, y puede valer entre 0° y 360° . Cuando va creciendo y llega a 360° , el valor N aumenta un entero y el $\Delta\Phi$, pasa a 0° ; y la inversa es cierta.

Pensemos que la de la λ de la L1 es de 20 cm y que al poder apreciar el $\Delta\Phi$ con precisión mejor del 1 %, la resolución interna es de orden sub-milimétrico.

Tenemos una incógnita, N , llamada **ambigüedad**. Realmente, la situación es similar a la que se tiene en el sistema de distanciometría electrónica por medición de fase. Pero al no poder modificar la frecuencia sobre la que medimos la fase, no es posible resolver la ambigüedad como se hace en distanciometría electrónica. La resolución de la ambigüedad



se hace en el proceso de cálculo, y no sólo de esta incógnita, sino también de las otras que tenemos, que son los estados de los relojes y por supuesto los tres incrementos de coordenadas entre receptores.

Debe destacarse que en el sistema es fundamental no perder el seguimiento de la fase para que la ambigüedad inicial no pueda variar. Si hay una pérdida de recepción por cualquier causa, la cuenta de ciclos se rompe y tenemos una pérdida de ciclos o *cycle slip*, talón de Aquiles del método, aunque mediante ajuste polinómico en post-proceso es posible restaurar la cuenta original y recuperar la ambigüedad inicial.

Aplicando este sistema de observación a dos estaciones podemos hacer las siguientes afirmaciones, refiriéndonos al posicionamiento relativo, que es el normalmente usado en las ciencias geográficas:

- El tratamiento de las ecuaciones generadas en la recepción instantánea común de un mismo satélite se llama **método de simples diferencias**, y minimiza o elimina errores de reloj de satélite.
- Las ecuaciones correspondientes a la recepción común en un momento dado de dos satélites en una posición en la órbita, **método de dobles diferencias**, facilitan la eliminación de las "cycle slips", y minimizan o eliminan los errores de reloj de satélites y de receptores, la indeterminación de órbitas y otras fuentes de error, porque al ser de magnitud similar, cuando estos errores son algebraicamente restados, tienden a cancelarse.
- Si el planteamiento matemático se hace tratando la recepción de dos satélites en una posición y luego en otra, **método de triples diferencias**, se eliminan los errores, igual que en las ecuaciones de dobles diferencias, pero se cancela la ambigüedad de ciclos.



Estos algoritmos se emplean automáticamente en el programa de cálculo.

La pérdida de ciclos puede ocurrir por muchas causas, desde el paso de un avión o pájaro, hasta por un relámpago o por disturbios ionosféricos, o a veces -raras-, sin causa aparente. Pensemos que, dada su lejanía y poco volumen físico, la fuente radiante, la antena del satélite, se puede considerar puntual, por lo que carece de "penumbra" radioeléctrica. Se comprende claramente la dificultad de trabajar en cercanías de árboles, tendidos eléctricos, estructuras, torres, etc.

BONDAD DE LAS OBSERVACIONES

En las técnicas GPS se consideran dos contribuciones diferentes al error esperable: **UERE** y **DOP**. Siguiendo la normalización propuesta por Paradisis y Wells, sus descripciones son:

UERE - User Equivalent Range Error (o error equivalente en la distancia por el usuario): es la contribución al error en la medida de la distancia producida por una sola fuente de error, suponiendo que la fuente de error no está correlacionada con las otras fuentes de error. Las fuentes individuales citadas son incertidumbres de efemérides, errores de propagación, marcha y estado de relojes y ruido de recepción. Con un mismo equipo, podemos tener por ejemplo un UERE desfavorable un día que haya tormentas eléctricas.

En un símil puramente didáctico, correspondería a la precisión con que mediríamos los ángulos en una triangulación, dependiente de la apreciación angular del instrumento que usáramos, del método de observación empleado, de la visibilidad existente, del observador, etc.

DOP - Dilution Of Precision (o dilución de precisión): es la contribución puramente geométrica a la incertidumbre de un posicionamiento. Es un valor aritmético descriptivo



de la "solidez" de la figura observable distanciométricamente, constituida por el receptor y los vectores que determinan con los satélites a la vista. Su valor ideal es 1, y si la geometría empeora el valor aumenta, llegándose a producir un **outage** o situación en la que, aunque haya sobrados satélites a la vista, deba suspenderse la observación porque el DOP llega a exceder de cierto valor preestablecido, como 6, por ejemplo, límite empleado habitualmente.

Realmente el DOP es un factor por el que debe ser multiplicado el error obtenido en las determinaciones de distancias a los satélites, para establecer el correspondiente error final de posicionamiento.

En la misma analogía mencionada antes de una triangulación clásica, correspondería a lo cercanos a 60° que fuesen los ángulos; cuanto más se separase el ángulo de 60° , mayor sería el DOP. En una trisección inversa, el "DOP" sería infinito si los vértices observados estuviesen contenidos en el círculo peligroso, que como se sabe es la situación en la que los vértices observados y el de estación están contenidos en una circunferencia, y no tiene solución por numerosos que fueran los vértices, dado que sus arcos capaces se superpondrían.

Los DOP más usados en nuestras técnicas son:

- **GDOP:** Tres coordenadas de posición y tiempo.
- **PDOP:** Tres coordenadas de posición.
- **HDOP:** Dos coordenadas de posición planimétrica.
- **VDOP:** Una coordenada de posición altimétrica.



PLANIFICACIÓN

La observación con un equipo GPS actual ha sido convertida por los avances técnicos en algo realmente anodino: en los equipos más evolucionados sólo hace falta estacionar la antena en el punto deseado y encender el receptor; a veces, ni siquiera eso, porque pueden programarse encendidos y apagados sucesivos. En cualquier caso, la desconexión final ultima la observación, sin más manipulaciones esenciales.

El observador se limita a comprobar que las condiciones de observación previstas se cumplen. El problema que sí se plantea es planificar el momento y la duración de la observación para obtener la calidad prevista, lo que depende del período elegido, de las condiciones de observación y de la configuración local de la constelación.

Cada vez que se realiza una observación de más de 12,5 minutos se adquiere un almanaque actualizado, que es una información expedita de órbitas, y permite, con el programa de ordenador adecuado, saber qué DOP habrá en determinado sitio y en determinado momento, y así planificar adecuadamente la observación.

MÉTODOS DE POSICIONAMIENTO

Como ya se mencionó en la descripción del sistema, el posicionamiento puede ser:

- absoluto o relativo.
- estático o dinámico.

Nos referiremos ahora al relativo, que es el usado en topografía y geodesia.

En el posicionamiento relativo siempre están involucrados varios equipos en recepción simultánea. De ellos, uno es el que establece el sistema local de referencia en el que



queremos determinar el vector definido por las dos antenas. Lo que se usa realmente no son las coordenadas que el sistema ofrece, sino los incrementos de coordenadas entre el receptor de referencia y el otro a posicionar, o sea la diferencia de posición entre ellos. Por ello, a este método se le puede denominar **diferencial** con total propiedad.

La gran ventaja de este método es que dos errores de posicionamiento muy similares o comunes en ambos puntos no tienen repercusión en los incrementos de coordenadas, porque se anulan mutuamente. Usando este método, los errores más difícilmente controlables (como el ionosférico) pueden ignorarse hasta en distancias medias. Por ello, en topografía, y siempre que se vayan a cubrir distancias de no más de 100 km, la característica bi-frecuencia (uso de la L2) y otras complejidades pueden evitarse, con el importante ahorro económico y de cálculo que ello supone.

Atendiendo al movimiento o no del receptor, el posicionamiento ofrece también otra alternativa: estático, cinemático y dinámico. Nos referiremos sólo a los dos primeros.

En el posicionamiento estático se hace la determinación, en tiempo diferido, de un conjunto único de coordenadas de un receptor a partir de la serie de observaciones realizadas. Pueden obtenerse soluciones **tan redundantes** como queramos, con sólo prolongar la observación. Las soluciones obtenidas pueden alcanzar precisiones muy altas.

Consiste en la utilización de un receptor base sobre un punto de coordenadas conocidas y otro receptor sobre el punto a medir. Ninguno de los dos receptores se mueve durante los prolongados tiempos de medición.



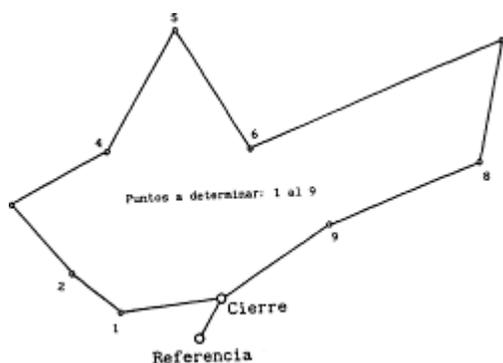
Es un método utilizado en geodesia para medir a largas distancias, y hoy por hoy es la manera más precisa para obtener coordenadas por GPS. Su precisión depende de los tiempos de medición y sobre todo del tipo de receptor empleado. Recordemos que este método se puede aplicar con receptores de fase de portadora L1 o con receptores de fase de portadoras en bi-frecuencia (L1+L2). En el caso de receptores de doble frecuencia la precisión del sistema viene a ser de 5 mm +1 ppm.

Las coordenadas medidas no son obtenidas por el usuario en el campo, sino que son calculadas en gabinete utilizando el software apropiado. Dicho software pone en relación las series de la estación (o estaciones de referencia) con las series de los receptores de medida. Como la estación de referencia ha estado ubicada en un punto de coordenadas conocidas, se puede saber en cada momento de la medición qué error aproximado estaban induciendo los satélites; dicho error es compensado sobre la serie del receptor medidor.

La idea principal de este método es que las señales que han llegado hasta la estación base han recorrido prácticamente la misma región atmosférica que las señales que han llegado hasta el receptor medidor, con lo cual ambas señales han estado sometidas al mismo tipo de degradaciones (sobre todo por efecto de la ionosfera).

Los receptores de dos portadoras (L1 y L2), al utilizar dos frecuencias distintas, permiten resolver mayor número de ambigüedades y dar mayor precisión; ello es así porque los retardos atmosféricos son función de la frecuencia de la señal. Luego, si utilizamos dos frecuencias distintas, podemos tener más información acerca de qué retardos y degradaciones ha tenido la señal en su camino hasta nuestro receptor.

El **posicionamiento cinemático** consiste en la determinación, en forma estática y relativa con respecto a un punto fijo, de grupos únicos de coordenadas, correspondiente cada uno



de ellos a un punto de los que configuran un itinerario cerrado. Por lo tanto, se determina un grupo de coordenadas por cada punto que haya. Como también se puede determinar un itinerario, puede denominarse con propiedad cinemático discreto.

La precisión obtenida es la correspondiente al sistema estático relativo de medición de fase de portadora, pero hay un drástico ahorro de tiempo.

Es necesario elegir dos puntos fundamentales: la referencia y el cierre. Con los demás puntos se configura un recorrido o itinerario que acceda a todos y tenga origen y final en el cierre.

Se usan dos receptores. Durante el trabajo, el fijo ("master") permanecerá continuamente estacionado en el punto referencia, mientras el móvil ("rover") recorrerá el itinerario proyectado, que empieza y termina en el punto cierre.

En este método es esencial que, desde el inicio hasta el final de la observación, ambos receptores realicen registros continuos de fase de portadora sobre un mínimo de 4 satélites de seguimiento común. La observación tiene dos partes claramente diferenciadas: **iniciación e itinerización**.



La observación de **iniciación** tiene un objetivo básico: para poder confrontar los resultados, hay que determinar unas constantes instrumentales para la pareja de receptores en funcionamiento, y sobre todo, las correspondientes ambigüedades de cada receptor con cada satélite seguido.

Hay tres posibles iniciaciones: base estática, base conocida, e intercambio; todo ello referido a las observaciones realizadas en los puntos referencia y cierre y a la base que establecen.

La segunda parte, la **itinerización** o **realización del itinerario**, es simple: como se sabe, en el punto referencia, y sin haber cesado en ningún momento el seguimiento, está el receptor fijo. En el punto cierre, y en idénticas condiciones, está el móvil. Entonces el móvil va recorriendo el itinerario con breves detenciones en cada punto (2 minutos, por ejemplo), hasta volver al cierre de donde partió.

El método ofrece un enorme potencial, aunque tiene fuertes limitaciones por la necesidad de no interponer obstáculos que provoquen pérdidas momentáneas de seguimiento de algún satélite.

Por supuesto, se pueden formar varias parejas de receptores, manteniendo común el "fijo" para varios "móviles", lo que aumenta espectacularmente el rendimiento.

Este método, poco explotado, es muy interesante en ciertos campos topográficos y cartográficos, pero el método pseudo cinematográfico, y sobre todo, el estático rápido, ofrecen alternativas eficaces.



ESTÁTICO RÁPIDO

La observación estática relativa es el fundamento de todas las observaciones de fase. Como se ha comentado en la descripción del sistema cinemático, hay una fase de iniciación en la que las observaciones se utilizan para resolver las incógnitas.

El método llamado "**estático rápido**" no es en absoluto un método de observación sino de tratamiento de las observaciones.

Potentes y complejos algoritmos de cálculo y comparación estadística, inicialmente desarrollados por Gerhard Beutler de la Universidad de Berna, permiten resolver todas las incógnitas con tiempos de observación notablemente reducidos.

La técnica se empezó a aplicar en receptores bi-frecuencia, idealmente con correlación del código P, lógicamente caros y sofisticados. Pero el poder correlacionar con el código P en ambas frecuencias aumentaba, por unidad de tiempo, el número de observaciones procesables.

Actualmente, la conversión del código P en el Y, secreto, impide su utilización directa y mengua la ventaja de los bi-frecuencia. Por otra parte, el incremento del número de satélites ha permitido a los mono-frecuencia el uso de estos algoritmos de cálculo, permitiendo tiempos de observación muy reducidos a los económicos receptores mono-frecuencia únicamente con código C/A.

COMBINACIONES

Las combinaciones posibles de las observaciones estáticas y dinámicas con el método cinemático y los sistemas absoluto y relativo ofrecen diversas combinaciones (menos cualquier combinación de cinemático con absoluto, que por su propio principio no puede existir).



Estático absoluto: para situar en tiempo inmediato una base en la Antártida o una patrulla militar en el campo de operaciones, por ejemplo.

Estático relativo: es el que ofrece la mayor precisión y es el plenamente utilizable en geodesia o topografía. Usando numerosos receptores se mejora la precisión, ya que una simple base carece de la configuración geométrica que tiene una red y que permite realizar ajustes. Siempre se disponen de los resultados en tiempo diferido.

Cinemático discreto, lógicamente relativo: ampliamente descrito. Precisión tan alta como en el estático relativo; resultados en tiempo diferido, pero con un notable ahorro de tiempo en la observación. Sólo utilizable en áreas no demasiado extensas y en condiciones favorables.

Dinámico absoluto: el usado por un navío en el océano, por ejemplo. Da su posición en tiempo real.

Dinámico relativo: hay varias opciones, diversificadas por las exigencias de precisión y de disponer de resultados en tiempo real, inmediato o diferido. Siempre involucra un mínimo de dos receptores, dado su carácter de relativo: uno fijo, en un punto de situación conocida, y otro móvil.

Si se usa el método de pseudo distancias, las indeterminaciones de órbitas, retardos y relojes provocan un aparente cambio de posición en la previamente conocida del fijo, o lo que es lo mismo, unos errores determinados en las distancias a los satélites.

El receptor fijo compara continuamente la posición aparente que le ofrece el sistema GPS con la real, y evalúa para cada instante su diferencia, que realmente es la corrección que ha de aplicarse a la posición aparente que el sistema GPS esté dando para el móvil si se



quiere conocer su posición real. Esto se llama **navegación diferencial**, que permite una navegación de considerable precisión, de orden métrico.

Si se necesitan resultados en tiempo real, como en la aproximación de una aeronave a una determinada zona, por ejemplo, es imprescindible una comunicación entre el receptor fijo y el móvil para que el primero transmita las correcciones al segundo. La comunicación se puede hacer por radio en distancias cortas o por vía satélite en grandes distancias; en este último caso, la precisión es lógicamente inferior al irse diferenciando los errores comunes a medida que aumenta la distancia.

Las correcciones obtenidas en el fijo también pueden almacenarse en memoria para luego ser aplicadas si la navegación no exige el posicionamiento en tiempo real, como podría suceder, por ejemplo, en un vehículo que estuviera haciendo inventarios de viales.

También puede usarse como una variante del cinemático si se observa diferencia de fase de portadora, que sería el **cinemático dinámico**, en el que no se determinan coordenadas de puntos en los que se detiene el receptor móvil, sino una trayectoria, no redundante, pero muchísimo más precisa que una navegación absoluta o bastante más que una diferencial con observación de pseudo distancias.

TRATAMIENTO DE LAS OBSERVACIONES

Tras observar en campo, se vuelca la información de todos los receptores utilizados en un ordenador provisto del programa adecuado, que siempre es facilitado por el fabricante. Es posible utilizar observaciones provenientes de receptores de diferentes marcas, usando el



formato normalizado "RINEX" (*Receiver INdependent EXchange Format*) al cual, y desde el cual, pueden convertirse los ficheros de observaciones.

Entonces se empiezan a procesar los datos correspondientes a cada vector observado cuya formación sea posible, cumpliendo la única condición de que la pareja de receptores que lo defina haya seguido satélites comunes (más de 3) en períodos comunes. Se emplean los mencionados algoritmos de simples, dobles y triples diferencias, y como resultado de todo este tratamiento se obtienen finalmente tres incrementos de coordenadas: AX, AY y AZ, correspondientes al vector (línea base) que une los dos centros radioeléctricos de las antenas de ambos receptores.

Tras determinar así el conjunto de todos los vectores observados, cuya escala y orientaciones absolutas son de precisión elevada, se termina lo que atañe al sistema GPS.

Los incrementos que definen estos vectores no son directamente utilizables, al ser observaciones "brutas", sin ajustar, y estar expresados en el sistema cartesiano geocéntrico ortogonal fijo a la Tierra, usado para establecer las coordenadas de los satélites, y con origen en el geocentro del sistema WGS 84, descrito cuando hablamos del sistema de referencia.

PASO AL SISTEMA LOCAL DE REFERENCIA

COMPENSACIÓN: La figura geométrica definida por estos vectores debe compensarse por mínimos cuadrados para que ajuste perfectamente. El cálculo se puede hacer en tridimensional, pero en cualquier caso, ofrece la misma garantía planimétrica que altimétrica. Tras la compensación disponemos de una nueva definición de cada vector,



pero formando una figura perfectamente ajustada. Recordemos que seguimos en un sistema geocéntrico cartesiano, con perfecta orientación y escala.

Los tres incrementos de coordenadas geocéntricas cartesianas que definen cada vector pueden convertirse en otros tres equivalentes, pero expresados en el sistema de referencia que necesitemos utilizar en el trabajo realizado: elipsoide definido, UTM, etc. También se pueden transformar en otros tres parámetros equivalentes, como acimut, distancia geométrica y ángulo de pendiente, por ejemplo.

Es evidente que si no se dispone de información sobre la ondulación local del geoide y su altura sobre el elipsoide, se obtendrán alturas sobre el elipsoide, pudiendo obtener altitudes erróneas.

TRANSFORMACIÓN: Si hay al menos tres puntos de esta figura tridimensional compensada de los que se conozcan las coordenadas en el sistema local de referencia que se vaya a usar, se puede hacer una transformación tridimensional de Helmert, con lo que se obtendrán coordenadas perfectamente utilizables y plenamente adaptadas al sistema local de referencia.

Normalmente se suele constreñir la escala y la rotación alrededor del eje Z (orientación), dado el carácter absoluto de estos parámetros en GPS.

Es de notar que la precisión obtenible con el posicionamiento relativo GPS ofrece normalmente precisiones superiores a las que tiene una red geodésica, obligado sistema de referencia local, por lo que al hacer la transformación de Helmert citada, nuestra figura GPS ya perfectamente ajustada probablemente "empeorará", se deformará en parte, ofreciendo residuos de transformación superiores a los de compensación. Esto es



inevitable al tener que adaptarse a unos puntos fijos dimensionalmente peor definidos que por el GPS.

También es importante resaltar de nuevo que la precisión altimétrica alcanzable con observaciones GPS es en todo idéntica a la planimétrica, siendo **nuestro** desconocimiento de la ondulación del geoide lo que puede causar incertidumbres en las altitudes. Al hacer la transformación antes mencionada, se adopta el plano de referencia del sistema geodésico local, por lo que la altimetría de los resultados GPS será tan buena como lo sea la de la red de referencia.

Para finalizar, sólo queda añadir que las elipses de error en una red observada con GPS normalmente son siempre "redonditas" y similares, tanto en el centro como en los bordes de la figura. Ciertamente, en los bordes de la figura son siempre algo mayores que en el centro, pero no varias veces mayores como en los sistemas clásicos.

2. Fundamentos de la Observación remota

Conceptos

La **Percepción Remota** (*Remote Sensing*) o **Teledetección** puede definirse como la ciencia y arte de obtener información de un objeto analizando los datos adquiridos mediante algún dispositivo que no está en contacto físico con dicho objeto.

El término Teledetección, traducción del inglés (*Remote Sensing*), también se define como el conjunto de conocimientos y técnicas utilizados para determinar características físicas y biológicas de objetos mediante mediciones a distancia, sin el contacto material con los mismos (Lasselin y Darteyre, 1991).



La percepción remota no engloba sólo los procesos que permiten obtener una imaginación desde el aire o el espacio, sino también su posterior tratamiento en el contexto de una determinada aplicación. Y se refiere no sólo a la captación de datos desde el aire o desde el espacio sino también a su posterior tratamiento. En otras palabras, la **Teledetección se describe como la técnica de adquisición y posterior tratamiento de datos de la superficie terrestre desde sensores instalados en plataformas espaciales, en virtud de la interacción electromagnética existente entre la tierra y el sensor, siendo la fuente de radiación ya sea proveniente del sol (Teledetección pasiva) o del propio sensor (Teledetección activa).**

La historia de la Percepción Remota comenzó hace unos 600 millones de años, cuando alguna forma inferior de vida animal diferenció algunas de sus células, volviéndolas fotosensibles. También durante millones de años dicho rudimento fotosensible evolucionó convirtiéndose en un poderoso y sofisticado sensor, el ojo humano. Este tuvo un imitador mecánico, la cámara fotográfica, que hizo su aparición hace algo más de un siglo y que fue muy mejorada durante la década de 1930 para ser aplicada a la fotografía aérea. La Segunda Guerra Mundial dio un gran impulso a la fotografía aérea así como a otras formas de percepción remota. Sin embargo, el "salto cuántico" en esta disciplina se produjo en la década de 1960 cuando las plataformas satelitales reemplazaron a las aéreas y los sensores electrónicos multi-espectrales, acoplados a computadoras, reemplazaron las cámaras fotográficas. El esquema operativo de un *satélite de observación* se representa en forma muy simplificada, en la figura 1.



Los objetos terrestres, iluminados por la radiación solar, reflejan ésta luego de introducir en ella modificaciones inducidas por la misma estructura y



composición de dichos objetos. La radiación reflejada es capturada por los sensores del satélite, siendo parcialmente procesada a bordo de éste y retransmitida a estaciones receptoras terrestres para su posterior procesamiento y análisis (*fuentes emisora, sensor, cerebro*).

Podemos considerar a la Teledetección y a la Fotogrametría como una misma técnica, ya que con las dos se obtienen los mismos productos, aunque las inercias de una y otra sean cualitativas y cuantitativas respectivamente.

Diferencias: Altura de Vuelo

Escalas

Tipo de información: Cuantitativa <> Cualitativa.

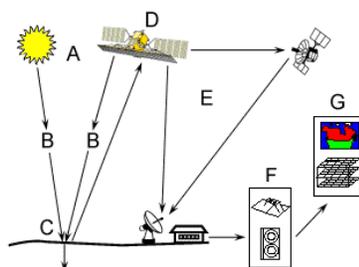
Semejanzas: La mayoría.

COMPONENTES DE UN SISTEMA DE PERCEPCIÓN REMOTA.

Los componentes básicos de un sistema de percepción remota son:

La escena, el sensor y el tratamiento de la información, con los siguientes elementos.

Los procesos que se involucran en la Percepción Remota son los siguientes:



A. Fuente de energía o iluminación: La Fuente de energía que supone el origen del flujo energético detectado por el sensor. Puede tratarse de un foco externo al sensor, en cuyo caso se habla de Teledetección pasiva, o de un haz energético emitido por este, conocido como Teledetección activa. La fuente de energía más importante es la energía solar.

B. Radiación y la atmósfera: Es el proceso en el que la energía electromagnética es propagada a través del espacio libre.

C. Interacción con la superficie terrestre: La cubierta terrestre, formada por distintas masas de vegetación, suelo, agua y construcciones, que reciben la señal energética procedente de la fuente de energía, y la reflejan o emiten de acuerdo con sus características físicas.

D. El sistema Sensor: Compuesto por el sensor propiamente dicho y la plataforma que lo sustenta. Tiene como misión captar la energía procedente de las cubiertas terrestres, codificarla y enviarla directamente al sistema de recepción.

E. Sistema de transmisión, Recepción, y proceso: El Sistema de recepción es en donde se recibe la información transmitida por el sensor; esta información es grabada en un formato apropiado y es distribuido a los intérpretes para su tratamiento visual o digital.

F. Interpretación y análisis: En esta etapa el intérprete analiza la información, donde es recibida la información transmitida por el sensor. Esta información es grabada en un



formato apropiado y es distribuido a los intérpretes para su tratamiento visual o digital esta información.

G. Aplicación: El último proceso es el de extraer la información necesaria para resolver un problema específico. Llega al usuario final, quien es el encargado de analizar el documento fruto de la interpretación, así como de dictaminar el impacto que de él se derive.

Radiación Electromagnética.

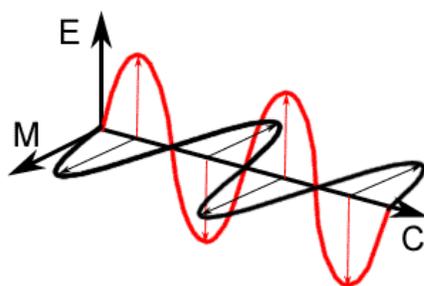
Anteriormente se definió la Teledetección como aquella técnica que nos permite obtener información a distancia de los objetos situados sobre la superficie terrestre. Para que esta observación remota sea posible, es preciso que entre los objetos y el sensor exista algún tipo de interacción.

Nuestros sentidos perciben un objeto sólo cuando pueden descifrar la información que este les envía. Por ejemplo, somos capaces de ver un árbol porque nuestros ojos reciben y traducen convenientemente una energía luminosa procedente del mismo. Esa señal, además, no es originada por el árbol, si no por un foco energético exterior que le ilumina. De ahí que no seamos capaces de percibir ese árbol en plena oscuridad. Este sencillo ejemplo nos sirve para introducir los tres principales elementos de cualquier sistema de Teledetección:

- Sensor (nuestro ojo),
- Objeto Observado (árbol)
- Flujo energético que permite poner a ambos en relación.

En el caso del ojo, ese flujo procede del objeto por reflexión de la luz solar. Podría también tratarse de un tipo de energía emitida por el propio objeto, o incluso por el sensor.

Éstas son, precisamente, las tres formas de adquirir información a partir de un sensor remoto:



- Por reflexión
- Por emisión
- Por emisión-reflexión

La primera de ellas es la forma más importante de Teledetección, pues se deriva directamente de la luz solar, la principal fuente de energía de nuestro planeta.

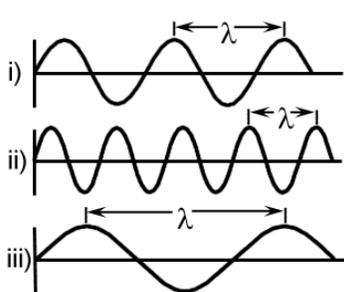
El sol ilumina la superficie terrestre, que a su vez refleja esa energía en función del tipo de cubierta presente sobre ella. Ese flujo reflejado se recoge por el sensor, que lo trasmite posteriormente a las estaciones receptoras. Entre superficie y sensor se interpone la atmósfera, que dispersa y absorbe parte de la señal original. De igual forma, la observación remota puede basarse en la energía emitida por la propias cubiertas, o en la que podríamos enviar desde un sensor que fuera capaz tanto de generar su propio flujo energético, como de recoger posteriormente su reflexión sobre la superficie terrestre. En cualquiera de estos casos, el flujo energético entre la cubierta terrestre y el sensor constituye una forma radiación electromagnética.

La radiación electromagnética tiene propiedades fundamentales y se comporta de varias maneras predecibles según los elementos esenciales de la teoría cuántica. Consiste en un campo eléctrico (E) que varía en magnitud y es perpendicular a la dirección en que la radiación está viajando, y un campo magnético (M) que está orientado en ángulos rectos al campo eléctrico. Ambos campos viajan a la velocidad de luz (c). Como es sabido, la energía se transfiere de un lugar a otro por tres procesos: convección, conducción y

radiación. Históricamente, las propiedades de la radiación electromagnética se han explicado por dos teorías aparentemente contrapuestas: aquella que la concibe como un haz ondulatorio (Huygens, Maxwell), y aquella otra que la considera como una sucesión de unidades discretas de energía, fotones o cuantos, con masa igual a cero (Plank, Einstein). Actualmente, parece que las dos teorías se pueden compaginar, pues se ha demostrado que la luz puede comportarse de acuerdo con ambos planteamientos. Según la teoría ondulatoria, la energía electromagnética se trasmite de un lugar a otro siguiendo un modelo armónico y continuo, a la velocidad de la luz, y conteniendo dos campos de fuerzas ortogonales entre sí: eléctrico y magnético. Las características de este flujo energético pueden describirse por dos elementos:

Longitud de Onda y Frecuencia

La primera hace referencia a la distancia entre dos picos sucesivos de una onda, mientras que la frecuencia designa el número de ciclos pasando por un punto fijo en una unidad de



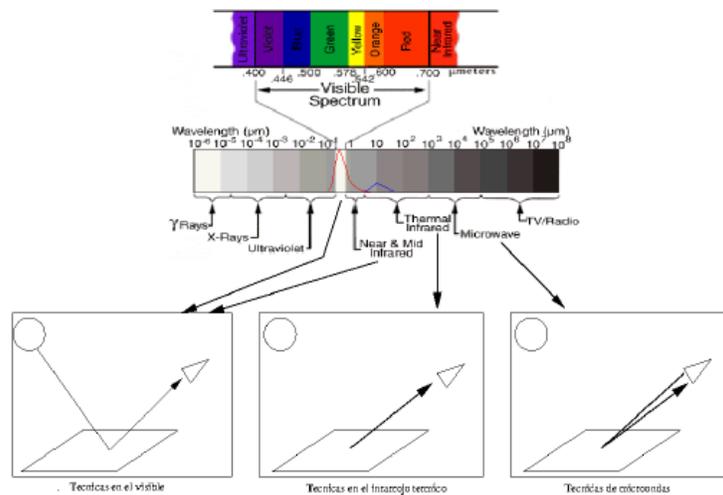
tiempo. Ambos elementos están inversamente relacionados: $c = \lambda F$, donde c indica la velocidad de la luz (3×10^8 m/s), λ expresa la longitud de onda (en micrómetros, $1 \text{ mm} = 10^{-6} \text{ m}$) y F la frecuencia (Hertzios, ciclos por segundo). En definitiva, a mayor longitud de onda, menor frecuencia y viceversa, por lo que basta con indicar un solo

término para identificar propiamente el tipo de energía mencionado. Gracias a la teoría cuántica, podemos calcular la cantidad de energía transportada por un fotón, siempre que se conozca su frecuencia: $Q = hF$, donde Q es la energía radiante de un fotón (en julios), F la frecuencia y h la constante de Planck (6.66×10^{-34} J.s). Sustituyendo en la ecuación $c = \lambda F$, podemos asimismo expresar: $Q = h(c/\lambda)$, lo que significa que a mayor longitud de onda o a

menor frecuencia, el contenido energético será menor y viceversa. Esto implica que la radiación en longitudes de onda largas es más difícil de detectar que aquella proveniente de longitudes cortas; de ahí que las primeras requieran medios de detección más refinados.

El Espectro Electromagnético.

Podemos definir totalmente cualquier tipo de energía dando su longitud de onda (o frecuencia).



El espectro electromagnético es la sucesión continua de esos valores de frecuencia, aunque conceptualmente se divide en bandas, en las que la radiación electromagnética manifiesta un comportamiento similar. La organización de estas bandas de longitudes de onda o frecuencia se denomina espectro electromagnético. Comprende desde longitudes de onda más cortas (rayos gamma, rayos X), hasta las kilométricas (telecomunicaciones). Las unidades de medida más comunes se relacionan con la longitud de onda. Para las más cortas se utilizan micrómetros, mientras que las más largas se miden en centímetros o



metros. Normalmente, a estas últimas (denominadas microondas) se las designa también por valores de frecuencia (en Giga hertzios= 10^9 Hz).

Región	Rango de longitudes
Rayos γ	$<0.03nm$
Rayos X	$0.03nm - 3nm$
Ultravioleta	$3nm - 0.3\mu$
Visible	$0.3\mu - 0.7\mu$
Infrarrojo	$0.7\mu - 300\mu$
Microondas	$300\mu - 20cm$
Radio-TV	$> 200cm$

Desde el punto de vista de la Teledetección, conviene destacar una serie de bandas espectrales, que son las más frecuentemente empleadas con la tecnología actual. Su denominación y amplitud varían según distintos autores, si bien la terminología más común es la siguiente:

Espectro Visible (0.4 a 0.7 mm).

Es la única radiación electromagnética perceptible por el ojo humano (de ahí su nombre). Coincide con la longitud de onda donde es máxima la radiación solar. Podemos localizar los distintos colores en las longitudes:

Suelen distinguirse tres bandas o colores elementales, que se denominan:

- **Azul:** 0.446 - 0.500 mm
- **Verde:** 0.500 - 0.60 mm
- **Rojo:** 0.600 - 0.70 mm

Los colores elementales están asociados a estas longitudes de onda.

Infrarrojo próximo (0.7 a 1.3 mm). A veces se le denomina también infrarrojo reflejado o fotográfico, puesto que parte de él puede detectarse a partir de filmes dotados de



emulsiones especiales. Resulta de especial importancia por su capacidad para discriminar masas vegetales y concentraciones de humedad.

Infrarrojo próximo (1.3 a 8 mm), en donde se entremezclan los procesos de reflexión de la luz solar y de emisión de la superficie terrestre. Resulta Idóneo para estimar contenido de humedad en la vegetación y detección de focos de alta temperatura.

Infrarrojo lejano o térmico (8 a 14 mm), que incluye la porción emisiva del espectro terrestre, en donde se detecta el calor proveniente de la mayor parte de las cubiertas terrestres.

Microondas (a partir de 1 mm), con gran interés por ser un tipo de energía bastante transparente a la cubierta nubosa. Se trata de las mayores longitudes de onda utilizadas en Teledetección. La observación de esta banda es compleja, y todavía no muy común. Pero el interés está creciendo debido a que en estas frecuencias la atmósfera es transparente y no existe el problema de que las cubiertas nubosas oculten la superficie terrestre.

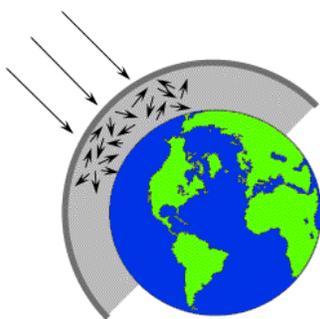
Como la emitancia de los objetos es muy baja en esta banda, es muy difícil detectar flujo. Los aparatos que lo hacen se llaman radiómetros de microondas. Por lo tanto, se construyen sensores para esta banda que emiten su propio haz de microondas, del que posteriormente recogerán su reflexión (Teledetección activa).

En este tipo de sistemas tiene mucha importancia la rugosidad de la superficie para efectos de cómo sea la señal de retorno. Por ejemplo, los suelos rugosos y secos tendrán

un mayor coeficiente de retro-difusión, y por tanto aparecerán en tonos más claros en las imágenes de radar.

Interacción de la Atmósfera con la Radiación Electromagnética.

Entre el sensor y la superficie terrestre se interpone la atmósfera, que interfiere de formas



diversas con el flujo radiante. Como es sabido, la atmósfera se compone de gases, anhídrido carbónico, oxígeno, ozono, nitrógeno y argón principalmente, vapor de agua y aerosoles.

La energía solar incidente en nuestro planeta está cifrada en 340 Wm^{-2} (vatios/m²). De toda ella, sólo 173 Wm^{-2} llegan a la superficie terrestre. Este factor de reducción o ALVEDO habrá

de ser tenido en cuenta a la hora de interpretar la respuesta de los objetos. En el trayecto seguido por la radiación electromagnética, ésta sufrirá tres fenómenos principales:

Absorción, Dispersión, Emisión.

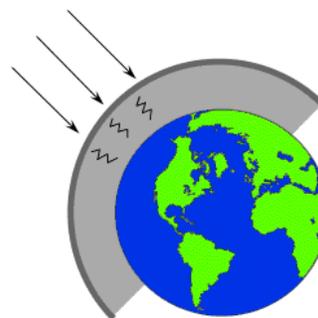
Sus efectos se manifiestan en un emborronamiento de la imagen; se reduce el contraste y la cantidad total de radiación que llega al sensor. Existen diversas correcciones en forma de algoritmos para tratar estos efectos.

Entre estos componentes, el anhídrido carbónico, el ozono y el vapor de agua son los principales responsables de la interacción con la energía electromagnética. Como ya se mencionó anteriormente, su efecto es triple:

- (i) **absorción:** proceso por el cual las moléculas y partículas de la atmósfera absorben la energía radiante (65 Wm^{-2}) y la transforman en energía interna que posteriormente será emitida en el infrarrojo térmico. La atmósfera reduce así la observación espacial a ciertas bandas concretas del espectro, llamadas ventanas atmosféricas. Pero si lo que se pretende es estudiar la atmósfera (satélites meteorológicos), los sensores se diseñarán para captar la radiación precisamente en aquellos lugares donde la absorción atmosférica sea más alta;
- (ii) **dispersión:** este fenómeno se traduce en un redireccionamiento o pequeño desvío del camino recto de propagación. Es causado por la interacción entre la radiación y los gases y partículas atmosféricas. La reflexión consiguiente a ese choque supone un aporte adicional a la radiancia proveniente de la superficie terrestre. Se reduce por tanto la radiancia directa, aumentando la difusa.
- (iii) **emisión:** como cualquier cuerpo con temperatura mayor a -273°K emite su propia radiación, esta lógicamente tiene mayor importancia en el infrarrojo térmico. Por tanto, su efecto es fundamental si pretendemos trabajar en dicha banda del infrarrojo térmico. Estos procesos introducen modificaciones, en ocasiones muy severas, en la radiación originalmente propagada entre la cubierta y el sensor.

Absorción atmosférica:

La atmósfera se comporta como un filtro selectivo a distintas longitudes de onda, de tal forma que en algunas bandas del espectro elimina prácticamente cualquier posibilidad de observación remota. Los principales causantes de esta absorción son:





- **Oxígeno atómico (O_2)**, que filtra las radiaciones ultravioleta por debajo de 0.1mm, así como pequeños sectores en el infrarrojo térmico y las microondas.
- **Ozono (O_3)**, responsable de la eliminación de la energía ultravioleta. inferior a 0.3mm, así como en un sector de las microondas (en torno a 27 mm).
- **Vapor de agua**, con una fuerte absorción en torno a 6 mm y otras menores entre 0.6 y 2mm.
- **Anhídrido carbónico (CO_2)**, que absorbe en el infrarrojo térmico (15mm), con importantes efectos en el infrarrojo medio, entre 2.5 y 4.5 mm.

Como consecuencia de esta absorción, la observación espacial se reduce a determinadas bandas del espectro, conocidas como ventanas atmosféricas, en donde la transmisividad de la atmósfera es suficientemente alta. Las principales ventanas atmosféricas son las siguientes:

- (i) espectro visible e infrarrojo cercano, situada entre 0.3 y 1.35 mm.
- (ii) infrarrojo medio: de 1.5 a 1.8 mm, 2.0 a 2.4 mm, 2.9 a 4.2mm. y 4.5 a 5.5 mm;
- (iii) infrarrojo térmico, entre 8 y 14 mm
- (iv) microondas, por encima de 20 mm, en donde la atmósfera es prácticamente transparente.

Estas ventanas atmosféricas son idóneas para realizar procesos de Teledetección, por lo que el diseño de los sensores espaciales tiende a ajustarse a estas bandas, evitando interferencias extrañas al fenómeno que pretende observarse, salvo la presencia de nubes, que absorben en todo el espectro óptico. Si se pretende, por el contrario, observar la atmósfera en lugar de la superficie terrestre, los sectores espectrales más convenientes



son, precisamente, aquellos en donde la absorción atmosférica es alta. Por esta razón, los satélites meteorológicos incorporan bandas de estas regiones del espectro.

Dispersión Atmosférica:

Es de gran importancia pero su efecto de absorción puede paliarse simplemente situando las bandas de observación en áreas donde la transmisividad atmosférica sea alta. Por ello, no resulta muy habitual encontrar ese problema en la interpretación de imágenes (salvo en caso de coberturas nubosas, normalmente evitables gracias a una buena selección de la imagen). Mucho más complejo, sin embargo, es soslayar el efecto de dispersión causado por la atmósfera, presente en mayor o menor grado en cualquier imagen adquirida con sensores remotos.

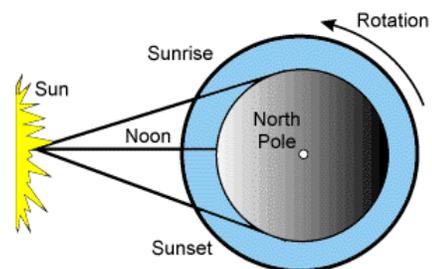
La dispersión de la radiación electromagnética es causada por interacción entre esta y los gases y partículas atmosféricas en suspensión. La reflexión consiguiente a ese choque supone un aporte adicional a la radiación proveniente de la superficie terrestre (luz atmosférica). En definitiva, se reduce la radiancia directa, aumentándose la difusa. Puesto que las partículas atmosféricas son muy variables en el tiempo y en el espacio, resulta muy complejo cuantificar su influencia final en la imagen adquirida por el sensor. Pese a ello, convendrá tener en cuenta este factor, especialmente cuando se pretenda convertir los valores digitales de la imagen a parámetros físicos, o cuando se aborden estudios multi-temporales.

Los principales causantes de la dispersión atmosférica son los aerosoles de vapor de agua. Los aerosoles son partículas en suspensión de origen muy diverso: oceánico, debidas al

movimiento de las aguas, o continental, polvo en suspensión o partículas emitidas por combustión. En función de su origen y características poseen muy variados tamaños, lo que implica distintos tipos de dispersión, ya que ésta es muy dependiente de su diámetro. De esta forma se habla de dispersión Rayleigh, cuando afecta a longitudes de onda inferiores al diámetro de la partícula; dispersión Mie, cuando se trata de partículas de diámetro similar a la longitud de onda, y de dispersión no selectiva, cuando se menciona las partículas de mayor tamaño.

Dispersión Rayleigh

Afecta a las longitudes de onda más cortas. Es la más conocida y la de mayor influencia en Teledetección, causante, por ejemplo, del color azul del cielo. En fotografía aérea, es muy claro su efecto por el tono azulado que presentan los fotogramas cuando se realizan desde cierta altura. Cualquier observador distante puede comprobar este resultado, especialmente en los meses de verano, cuando se difumina el paisaje al divisarlo a largas distancias.



7. Dispersión Mie:

También depende de la longitud de onda, si bien en menor grado que la anterior. Los aerosoles y polvo atmosférico son los principales responsables de este tipo de dispersión, aunque también está presente en incendios forestales o en brumas costeras.

Dispersión No Selectiva



Afecta por igual a diversas longitudes de onda. Por esta razón, las nubes o nieblas tienden a aparecer blancas, ya que dispersan por igual toda la luz visible. Los procesos de dispersión son muy complejos y difíciles de cuantificar en la imagen resultante. Normalmente no se dispone de datos coetáneos a la adquisición de ésta, por lo que la corrección atmosférica se basa en relaciones entre elementos de la propia imagen. Esta estimación puede ser burda en algunos casos; dificulta, en última instancia, el uso de las imágenes adquiridas por sensores espaciales como medida cuantitativa de parámetros biofísicos.

Emisión atmosférica:

El efecto de emisión atmosférica resulta fundamental en el trabajo dentro del infrarrojo térmico si se pretenden obtener mediciones de temperatura a partir de las imágenes espaciales. Al igual que cualquier cuerpo por encima del cero absoluto, la atmósfera emite energía calórica, por lo que ese parámetro debe considerarse por separarlo de la emitancia espectral proveniente del suelo. Esos procesos de corrección atmosférica se han abordado habitualmente mediante el análisis combinado de la señal captada en dos bandas situadas en el infrarrojo térmico. Son los denominados algoritmos de ventana partida (*split window*).

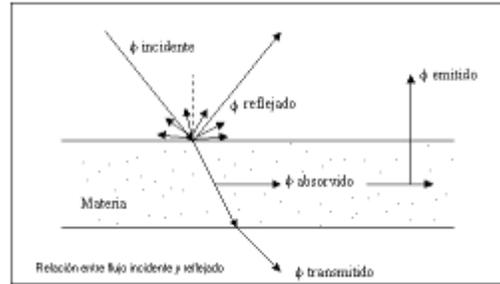
Interacción de la Radiación con las Superficies:

Se denomina dominio óptico del espectro a aquel grupo de longitudes de onda directamente dependientes de la energía solar. Se analiza



I=Energía Incidente
A=Energía absorbida
T=Energía transmitida
R=Energía reflejada

además lo correspondiente al infrarrojo medio cuando



los procesos de reflexión son dominantes sobre los de emisión. En consecuencia, conviene analizar con más detenimiento en este apartado las características espectrales del sol, fuente principal de energía radiante en nuestro planeta. Más adelante se analizará el comportamiento de las principales coberturas terrestres frente a este tipo de radiación.

El sol se encuentra a una temperatura radiante próxima a los 6.000 K. Esto implica, según la ley de Wien, que su máxima emitancia espectral se produce en torno a los 0.45 μm , coincidente con el color verde apreciado por nuestros ojos. La curva espectral de la radiación solar se asemeja bastante a la de un cuerpo negro a esa temperatura.

Ya hemos visto antes que esta banda constituye el grupo de longitudes de onda directamente dependientes de la luz solar. Es la banda de mayor interés para la observación remota de la superficie terrestre.

La radiación, cuando entra en contacto con la superficie terrestre, puede sufrir cuatro procesos diferentes:



- Parte será reflejada con un ángulo similar a aquél con el que haya incidido. Esta reflexión recibe el nombre de **ESPECULAR**.
- Parte será reflejada uniformemente en todas direcciones. Reflexión **LAMBERTIANA**.
- Parte será absorbida por el objeto, de modo que éste sufra un calentamiento. Más adelante perderá esa energía en forma de radiación en el infrarrojo térmico; el cuerpo se enfriará.
- Parte será transmitida a otros objetos o a la superficie terrestre y no volverá al espacio exterior, donde se encuentra el sensor.

El flujo radiante se descompone en tres términos:

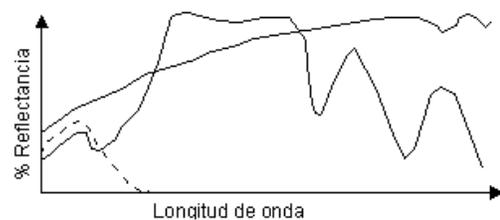
$$f_i = f_r + f_a + f_t$$

Flujo incidente es igual al reflejado más el absorbido más el transmitido.

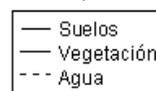
Dividiendo todos los términos entre f_i obtenemos: $1 = r + a + t$.

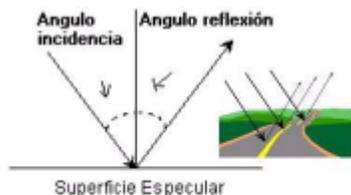
Siendo rigurosos, dependen de la longitud de onda: $1 = r_l + a_l + t_l$

El hecho de que esta proporción $r_l + a_l + t_l$ varíe con la longitud de onda es interesante para poder distinguir diversas zonas, conociendo precisamente esa variación de proporciones. Las divergencias espectrales entre superficies permitirán distinguirlas (especialmente las referidas a reflectividad, que suponen la principal fuente de información en Teledetección), una vez hayan sido corregidos los efectos causados por la atmósfera y otras anomalías. En el espectro visible, esta divergencia se manifiesta en lo que llamamos 'color'.



Firmas espectrales típicas para tres tipos de superficies





Superficie Especular:
Es aquella superficie ideal en la que se cumple perfectamente la ley de la reflexión (Angulo incidencia=Angulo reflejado)

Teniendo en cuenta el componente de reflectividad (r_1), que es lo que podemos medir con el sensor, podremos saber qué tipo de superficie es la que da esa firma espectral.

En cuanto a las condiciones de observación, conviene tener presente que la cantidad de energía que llega del sensor depende del ángulo

con que la superficie refleje la energía incidente, así como del que formen el haz incidente con la posición del sensor.

Esta geometría de observación está estrechamente ligada a la rugosidad que presenta la superficie. A este respecto, pueden distinguirse dos tipos de cubiertas: aquellas que reflejan la energía con el mismo ángulo de flujo incidente (**ESPECULAR**) y aquellas que lo reflejan uniformemente en todas las direcciones (**LAMBERTIANA**). En el primer caso, el sensor sólo recibe energía reflejada del suelo si está situado en la dirección del ángulo de reflexión, siendo esta nula en cualquier otro caso. Si la superficie es lambertiana, la radiancia es constante en cualquier ángulo de reflexión.

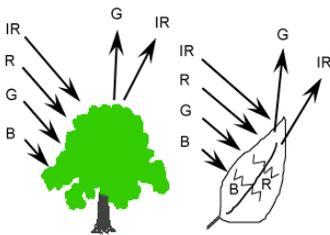


Superficie Lambertiana
SUPERFICIE LAMBERTIANA:

Es aquella superficie, también ideal, que refleja la radiación incidente de manera uniforme en todas las direcciones.

Comportamiento espectral teórico para los tres elementos fundamentales del paisaje en el dominio óptico:

Vegetación vigorosa



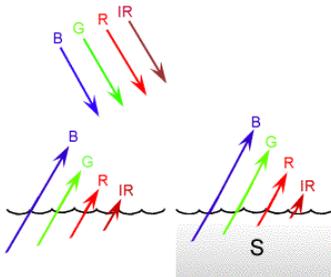
Es especialmente variable, dependiendo de múltiples factores: estado fenológico, forma y contenido de humedad. De manera general se puede decir que presenta una reducida reflectividad en las bandas visibles, con un máximo relativo en la región del verde (0.55 μm). Esto es debido al efecto absorbente de los pigmentos

fotosintéticos de las hojas. De entre estos, las hojas contienen clorofila, que absorbe las longitudes de onda del rojo y azul. Por eso, en vegetación no vigorosa la clorofila no absorbe tanto y la hoja se torna de un color amarillento (verde + rojo).

La estructura celular interna da la elevada reflectividad en el infrarrojo cercano, que luego va reduciéndose paulatinamente hacia el infrarrojo medio. Por tanto, el contraste más nítido se presenta entre las bandas visibles (especialmente rojo) y el infrarrojo cercano. Cuanto mayor sea ese contraste, mayor será el vigor de la vegetación.

Estos tres parámetros: cantidad de pigmentos, estructura celular y contenido en agua, manifestados por la respuesta espectral de esa masa de vegetación de la que forman parte, puede servir para discernir entre unas especies y otras, su nivel de desarrollo, e incluso entre su estado sanitario.

Agua:



Absorbe o transmite la mayor parte de la radiación visible que recibe. Por lo tanto, presentará una curva espectral plana, baja y de sentido descendente. De todas formas, en aguas poco profundas, la reflectividad aumenta. Los factores que afectan este comportamiento son: profundidad, contenido en materias en suspensión (clorofila, arcillas y nutrientes) y rugosidad de la superficie (éste último factor extremadamente importante). Por su parte, la nieve ofrece un comportamiento completamente diferente al del agua, con una reflectividad elevada en las bandas visibles, reduciéndose drásticamente en el infrarrojo cercano.

Suelos:

Muchas veces son invisibles, debido a la vegetación.

El comportamiento espectral del suelo desnudo es mucho más uniforme que el de la vegetación, mostrando una curva espectral bastante plana y de carácter ascendente. Los principales factores que intervienen en este caso son la composición química del suelo, su textura y estructura, y el contenido de humedad.

Por ejemplo, un suelo de origen calcáreo tiende al color blanco, indicando alta reflectividad en todas las bandas visibles, mientras que los arcillosos ofrecen una mayor reflectividad en el rojo, como consecuencia de su contenido de óxido de hierro. La reflectividad espectral presenta mayores valores en suelos de textura gruesa, apelmazados, secos y sin materia orgánica. Como ya se ha dicho para la vegetación, el contenido de humedad es uno de los elementos destacados en la reflectividad en



longitudes de onda largas (infrarrojo medio), y así, los suelos con alto contenido de humedad darán una reflectividad baja en esa banda. Cuanto más seco y apelmazado sea un suelo, mayor será su reflectividad.

Comportamiento espectral teórico para los tres elementos fundamentales del paisaje, en el dominio del infrarrojo térmico

Características espectrales de los suelos

Los suelos están afectados por muchos componentes, lo que hace que su respuesta espectral sea muy amplia y poco tipificable. Además, cuando la información proceda de una imagen digital, el tamaño del píxel condiciona la heterogeneidad de la información contenida en dicha celda.

Centrándonos en los aspectos más importantes, abordaremos el estudio en cuanto a textura, humedad, contenido de materia orgánica, composición mineralógica y rugosidad de la superficie. Debe tenerse presente el hecho de que muchas veces el suelo va a estar situado bajo una cubierta vegetal, por lo que la respuesta que obtenga el sensor en este caso es una situación especial que no puede extrapolarse de lo que a continuación se comenta. En cualquier caso, por medidas indirectas, y por anomalías en la capa de vegetal, puede obtenerse información litológica del suelo.

Comportamiento espectral de las coberturas vegetales.

Es fácil admitir que el comportamiento espectral de las hojas, de forma individual, es importante para comprender las características espectrales de las coberturas espectrales, pero no permite explicarlas por completo.



En condiciones naturales, las cubiertas vegetales se componen de varias capas de hojas. La superior producirá sombras y/o enmascaramientos espectrales sobre las inferiores, dándose una reflectancia global resultado de la combinación de sombras y superficie foliar. La cantidad de superficie sombreada dependerá de factores como el ángulo de iluminación, el tamaño, la forma y orientación de las hojas, etc.

El sombreado tiende a disminuir la reflectancia global con respecto a la observada para hojas individuales, si bien esta disminución es mucho menor en el NIR que en el visible. Esto se justifica por el hecho de que gran parte de la radiación infrarroja se transmite a través de las capas más altas de la vegetación, siendo parcialmente reflejada por las capas inferiores y transmitida de nuevo a través de las hojas superiores. Se obtiene así como resultado un incremento relativo de la reflectancia global en el NIR. Por el contrario, la radiación visible es absorbida en proporciones mucho mayores.

Un fenómeno muy interesante es el que Collins denominó *red shift* (desplazamiento del rojo), que se basa en el cambio que se produce en la respuesta espectral de los cultivos según se aproximan a la madurez fisiológica. Así, la zona correspondiente al espectro del rojo lejano (aproximadamente $0.74\mu\text{m}$.), donde disminuye la absorción de clorofila y aumenta la reflexión infrarroja, experimenta un desplazamiento hacia regiones de mayor longitud de onda a medida que los cultivos se aproximan a su madurez. A partir de ello, no solo es factible diferenciar el estado vegetativo de un cultivo, sino discriminar entre distintas especies vegetales.

Como nota importante está el hecho de que, en aplicaciones de discriminación agrícola, es conveniente utilizar información relativa a la respuesta espectral de cada una de las



distintas épocas de la campaña. Con ello se puede conocer la evolución seguida a lo largo del tiempo y evaluar las variaciones estacionales en el crecimiento y vigor de las plantas.

Sistemas espaciales de Teledetección.

Históricamente fueron la placa o el film fotográfico sobre plataformas aéreas los sensores remotos por excelencia, pero a partir de las últimas décadas debieron convivir con los sensores electrónicos, cada vez más numerosos y sofisticados, montados sobre plataformas suborbitales y satelitales.

La Teledetección desde plataformas aéreas, sea con placas, lms fotográficos o sensores electrónicos, si bien sigue constituyendo una muy valiosa fuente de información, a partir de 1960 debió comenzar a compartir su rango de aplicaciones con las tecnologías de Teledetección desde plataformas satelitales.

En efecto, hasta 1946 la percepción remota se efectuaba fundamentalmente desde aviones o globos. En 1946 se tomaron las primeras fotografías desde cohetes V-2 capturados a los alemanes, y tales experiencias fueron decisivas para ilustrar el valor potencial de la fotografía desde alturas orbitales. Dicho potencial se volvió más aparente a través de las misiones espaciales a partir de 1961: Mercury, Gemini y Apolo.

El Apolo 9, en particular, realizó la primera experiencia de *fotografía orbital multi-espectral* para el estudio de los recursos terrestres. Estas experiencias fueron continuadas en misiones posteriores (Skylab, Apolo-Soyuz Test Project, etc.). Sin embargo, el año 1972 marca un hito en la carrera espacial: fue en dicho año que EE.UU. lanzó en un vehículo Delta el primero de una serie de *satélites ópticos* especialmente destinados a monitorear



recursos terrestres. Dicha serie, que se denominó ERTS (*Earth Resources Technology Satellites*) fue algo re-denominado LANDSAT, de la cual aún permanecen operativos el LANDSAT 5 y el LANDSAT 7, este último recientemente puesto en órbita. A su vez Francia, con la participación de Suecia y Bélgica, lanzó en 1986 en un vehículo Ariane el primer satélite de la serie SPOT (*Système Pour l'Observation de la Terre*) continuando operativos el SPOT 2 y el SPOT 4.

Actualmente existen numerosas plataformas satelitales similares al LANDSAT y al SPOT pertenecientes a diferentes países (EE.UU., India, Rusia, Japón, etc.). También existen otros sistemas satelitales destinados sobre todos a estudios meteorológicos y ambientales como por ejemplo los de la serie NOAA (*National Oceanic and Atmospheric Administration*) y los meteorológicos geostacionarios como los de la serie GOES (*Geostationary Operational Environmental Satellite*) y METEOSAT. Sería muy extenso referirnos aquí a la diversidad de plataformas orbitales de observación actualmente en órbita, pero podemos decir que el desarrollo que se inició en la década de 1960 se ha vuelto vertiginoso. Distintos factores han condicionado este comportamiento, lo que ha llevado a la Percepción Remota a su *etapa verdaderamente comercial*. Uno de tales factores fue la liberación para usos civiles, particularmente luego del fin de la Guerra Fría, de tecnología reservada hasta entonces para uso militar.

Debe señalarse que los sistemas satelitales mencionados pueden registrar las imágenes terrestres simultáneamente en varias bandas espectrales, usualmente ubicadas en las regiones *visibles e infrarrojo cercano, infrarrojo medio e infrarrojo termal* del espectro de las radiaciones electromagnéticas. Esto permite discriminar la naturaleza de los objetos terrestres e incluso clasificarlos. Los sistemas *óptico-telescópicos* de estos satélites enfocan las escenas terrestres sobre arreglos de *detectores de estado sólido* sensibles a las



regiones espectrales para las que se desea registrar las imágenes. Las *señales analógicas* generadas por los detectores son *digitalizadas* a bordo del satélite para su retransmisión a estaciones rastreadoras terrestres.

El primer satélite de observación terrestre fue lanzado en 1960. Se trataba del **TIROS** (*Television and Infrared Observation Satellite*), un prototipo operacional de una serie de satélites meteorológicos experimentales para el seguimiento y estudio de los frentes nubosos. Sucesores de aquella serie serán los actuales NOAA–AVHRR, que hoy proporcionan datos meteorológicos para previsiones diarias del tiempo en gran parte del mundo.

En 1973, fruto del interés iniciado en la década anterior, se lanza el **LANDSAT 1** (*Land Satellite 1*), que supondrá el primer satélite diseñado específicamente para la observación sistemática de las zonas terrestres del planeta.

De ahí a la actualidad, el interés se ha materializado con nuevos satélites dedicados al seguimiento de recursos terrestres en los que se pretende, en principio, una mayor resolución espacial, en detrimento de la resolución espectral. Es el caso concreto del programa SPOT y sus derivados IRS, Orbimage, IKONOS, etc.

De forma paralela, los últimos indicios indican que los sensores híper-espectrales en satélites, poco a poco van a ir tomando el mismo cariz que tienen los actuales sensores híper-banda aerotransportados. Por tanto, en breve será posible contar con este tipo de información a nivel sistemático para todo el planeta.



Tipos de sistemas

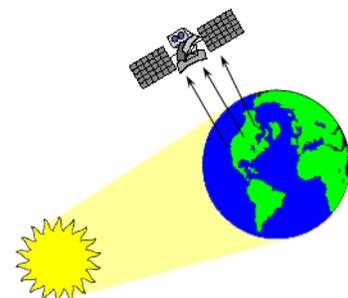
Al abordar el estudio de las plataformas satelitales, sobre todo de aquellas destinadas a la exploración de recursos terrestres, el objeto principal de nuestro estudio va a recaer sobre la tipología del sensor que incorpora. Esto es así por el hecho de que las aplicaciones a las que se pueda destinar un satélite van a venir marcadas directamente por la resolución (ya sea radiométrica, espectral, espacial o temporal) que el sensor pueda acometer.

Este condicionante, junto con la resolución temporal, marca definitivamente las características propias de cada plataforma orbital. Por lo tanto, previamente al estudio pormenorizado de cada satélite, se precisa tener una idea básica del funcionamiento de los diferentes sensores incorporados. Paralelamente, dado que la elección del tipo de sensor se realiza en el momento de diseño del satélite, y conocido aquel, nos podemos hacer una idea de los objetivos pretendidos por el constructor al abordar su construcción.

Un sistema de percepción remota o Teledetección requiere por los menos de tres componentes: **foco energético, superficie terrestre y sensor**. Entre las diversas formas para clasificar a los sensores remotos, una de las más habituales es mediante el procedimiento para recibir la energía procedente de las distintas cubiertas.

Sensores Pasivos

Recogen la energía electromagnética procedente de las cubiertas terrestres, tanto la reflejada por los rayos solares (visible e IR) en su iteración con la superficie, como la emitida por la propia temperatura terrestre (MIR).





Estos sensores, en función de la técnica utilizada para recibir la radiación, se pueden dividir en **fotográficos, óptico-electrónicos y radiómetros de microondas**. Los sensores fotográficos están basados en cámaras fotográficas, al estilo de las utilizadas en fotogrametría aérea, pero situadas en estaciones orbitales. Su funcionamiento, por tanto, se basa en impresionar una emulsión fotosensible recogiendo de esta forma el objeto enfocado por la parte óptica.

Cámara CCD o cámaras de Vidicon: su aplicación fue muy extendida en los primeros satélites, y aún hoy, se siguen utilizando aplicados a proyectos de gran escala, o en calibración de imágenes tomadas por otros sensores.

Los sensores óptico-electrónicos son una derivación de los fotográficos. Presentan una óptica más o menos similar a la fotográfica pero con un sistema de detección electrónica que evita la dependencia de una superficie sólida como emulsión.

Se dividen en Exploradores de Barrido (escáneres) y Exploradores de Empuje (*pushbroom*).

- 1) En los **exploradores de barrido**, un espejo móvil oscila de forma perpendicular a la trayectoria del satélite, lo que permite recoger información terrestre a ambos lados de su nadir, pero de forma individual para cada punto (píxel) de la línea explorada.
- 2) En los **exploradores de empuje** se elimina el espejo oscilante y se incorpora una cadena de detectores. Con ello se consigue que de una sola vez se recoja la línea completa que conforma el campo de visión perpendicular a la traza del satélite. El sensor HRV de SPOT es de este tipo. Así, con su 'viga' de 6000 detectores, cada uno con 10m. de resolución en pancromático, puede tomar una línea de 60.000m (60km.) de traza, de una sola vez.



De lo anterior, es fácil deducir los pros y contras de cada sistema. Así, los exploradores de barrido, al escanear píxel a píxel, presentan un mayor problema en el movimiento del espejo que puede desconfigurarse y/o presentar problemas de calibración geométrica. Asimismo son más lentos en la toma de una sola línea. En cambio, la calibración radiométrica entre píxeles de la imagen es casi inexistente, además de tener una mayor sensibilidad espectral (0,4-12,6 μ m).

Los exploradores de empuje son capaces de lograr una mayor resolución espacial al eliminar la parte móvil, pero a costa de ello se les añade una dificultosa calibración entre los sensores. Su rango espectral está comprendido entre el visible y el NIR.

Si los sensores anteriores estaban destinados más o menos a un mismo rango espectral, los radiómetros de microondas se salen de este intervalo y se basan en longitudes de onda larga, entre 1 y 100mm.

A su favor está que no los afectan las situaciones atmosféricas y las condiciones de iluminación de la escena. Pero tienen en contra su pobre resolución espacial.

Sensores de Antena o Radiómetros de Microondas: Tienen poca resolución espacial. Permiten cartografiar zonas heladas.

Sensores Activos

Se basan en la emisión de un haz de energía que posteriormente recogen tras su reflexión en la superficie de estudio. En función del comportamiento del haz sobre la superficie (difusión, reflexión, absorción), es posible estimar las características (rugosidad, humedad, etc.) y, por lo tanto, el tipo de cubierta



El **RADAR** (*Radio Detection And Ranging*) es el sensor activo por excelencia. Trabaja en el rango de 1mm-1m. Por ello, presenta caracteres similares al microondas, es decir, independencia de condiciones atmosféricas e iluminación, pero también baja resolución espacial.

La gran apuesta del radar se sitúa en el sistema aerotransportado, que permite aumentar la resolución al disminuir la altura. Por otro lado, gracias al sistema SAR, también se está consiguiendo una mayor resolución espacial basada en un alargamiento ficticio de la antena receptora.

En cualquier caso, la problemática se presenta por deformaciones geométricas de carácter lateral y también por deformaciones propias del sistema debido a la distancia variable entre la superficie de estudio y el emisor, el efecto de relieve inverso, etc.

Debido a su flexibilidad, puede trabajar en cualquier tipo de condición atmosférica. Por ello, el radar ha acaparado la atención de la comunidad científica internacional, especialmente en países ecuatoriales, donde son persistentes las masas nubosas. Es bastante utilizado en aplicaciones militares, así como en el control del tráfico aéreo. El radar trabaja en la región de las microondas.

El **LIDAR** (*Light Detection And Ranging*) es otro de los sensores activos. Dadas sus características intrínsecas, sus aplicaciones están muy diferenciadas de las de los otros sensores, incluso del radar, ya que la menor longitud de onda del **LIDAR** permite mayor detalle en la observación, si bien no penetra en las estructuras nubosas.

Su metodología es la de un telescopio que envía pulsos de luz polarizada entre el UV y el NIR. Estos pulsos se pueden enviar a satélites orbitales (serie LAGEOS) que, a modo de

reflectores, la devuelven a la Tierra, de manera tal que se pueden medir distancias como si se tratara de un distanciómetro cualquiera.

Sus precisiones son centimétricas y es de gran utilidad en geodesia para coordenadas de redes geodésicas, deriva continental, etc.

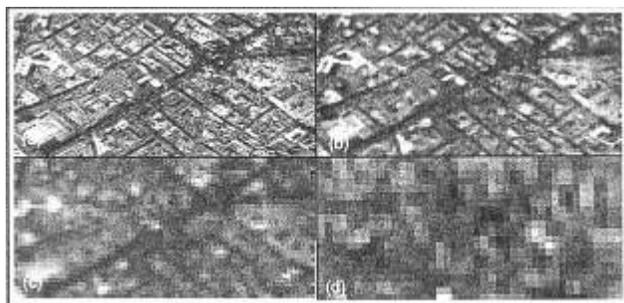
Sin necesidad de satélites, estos pulsos de luz inducida también se están utilizando haciéndolos interactuar con la atmósfera y consiguiendo estudios de composición del aire (contaminación atmosférica, aerosoles, medición de humedad, velocidad del viento, etc.).

Desde plataformas aéreas, a través de fluorescencia inducida de los rayos UV, permite diferenciar manchas de petróleo, humedad en la vegetación, biomasa, porcentaje de pigmentos, etc.

En los últimos años, los sensores **LIDAR** están tomando especial auge aplicados de forma efectiva para medir alturas de objetos, por ejemplo la altura relativa de una cubierta vegetal con respecto a la superficie del suelo, o la profundidad de una capa de agua.

Resolución de un sistema sensor.

Una característica muy importante de un sensor es la resolución. La resolución de un sensor es la habilidad para discriminar información de detalle. Existen cuatro tipos de resolución que se puede medir:



Resolución espacial: determina el objeto más pequeño que puede ser distinguido en una imagen. El píxel es la unidad mínima de información que contiene una imagen, por lo cual la resolución espacial es la equivalencia existente entre el píxel y su dimensión. Aquí mostramos un ejemplo de un

sensor con varias resoluciones espaciales sobre una misma zona:



- a) Resolución de 1m
- b) Resolución de 5m
- c) Resolución de 10m
- d) Resolución de 30m

Resolución espectral: determina el número y las anchuras de las bandas espectrales que puede discriminar el sensor.

Resolución radiométrica: es la capacidad para detectar variaciones en la radiancia espectral que recibe, y cuanto mayor sea, mejor se podrá interpretar la imagen.

Resolución temporal: se refiere a la frecuencia de cobertura que proporciona el sensor, es decir, la periodicidad con la que adquiere imágenes. Los sensores meteorológicos suelen obtener imágenes cada 30 minutos.

LANDSAT:

Esta familia de satélites tuvo su inicio en 1972, cuando la agencia espacial norteamericana decidió invertir esfuerzos en la observación espacial de recursos terrestres. Su buena acogida fue consecuencia de una resolución excepcional para la época. La cobertura sistemática global junto con las características geométricas, radiométricas y temporales, la hacían ideal para estudios con un sinfín de posibilidades.

Las aplicaciones han ido aumentando a lo largo de la vida del satélite. Se iniciaron con el ánimo de elaborar un inventario agronómico, previsión de cosechas, evaluación de zonas irrigables, etc., y se han ido extendiendo a diferentes y muy variados campos: previsión de incendios, cálculo de biomasa, cartografía litoral, etc.



Los tres primeros satélites Landsat tenían una estructura similar, por lo que se suelen estudiar de forma conjunta. En principio incorporaban el sensor MSS (*Multispectral Scanner*) a los que se acoplaban un conjunto de tres cámaras de vídeo denominadas RBV (*Return Beam Vidicon*). Estas cámaras no dieron un resultado satisfactorio y en los Landsat 4 y 5 fueron sustituidas por el sensor TM (*Thematic Mapper*).

De esta forma, los satélites 4 y 5 tienen un doble sistema de exploradores por barrido, que más que completarse, se superponen. La idea fue dejar los MSS con el fin de dar continuidad a las series de datos efectuadas con Landsat 1, 2 y 3.

	Landsat 1-3	Landsat 4-6	Landsat 7
Altitud	907-915 km.	705 km.	705 km.
Inclinación	99.2°	98.2°	98.2°
Orbita	polar, Solar-sincrónica	Polar, Solar-sincrónica	Polar, Solar-sincrónica
Paso por el Ecuador	09h30 AM	09h30 AM	10h00 AM
Periodo de Revolución	103 m	99 m	99 m

El lanzamiento del Landsat 6 en 1996 resultó un fiasco, por lo que la serie ha continuado con el

Landsat 7, lanzado el 15 de abril de 1999, y cuya primera imagen fue proporcionada 3 días después.

Landsat 7

Anchura de Barrido:	185 kilómetros
Altitud:	705 kilómetros
Quantización:	8 bits
Capacidad de almacenamiento a bordo:	~375 Gb
Inclinación:	Solar-sincrónica, 98.2 grados
Paso Ecuatorial	Descendente; 10:00am +/- 15 min.
Vehículo de Lanzamiento:	Delta II
Fecha de Lanzamiento:	Abril 1999

Ya se ha prescindido definitivamente del sensor MSS, quedando únicamente el TM, dado que dispone de una mayor resolución tanto espacial como espectral. Asimismo, se añaden



nuevas características con el fin de adecuar el satélite a los tiempos actuales y poder ofrecer a los usuarios aquellas imágenes que éstos demandan.

En un principio la denominación para el sensor a bordo del Landsat 7 iba a ser ETM (*Enhanced*

Thematic Mapper), es decir, *Mapeador Temático Realzado*, pero finalmente a este nombre se le ha añadido un Plus final, ETM+ (*Enhanced Thematic Mapper Plus*), con intenciones de conferirle al nombre un carácter más innovador.

Las características del ETM+ coinciden con las del Landsat 4 y 5, excepto que se añade una banda extra pancromática de 15m de resolución, que está co-referenciada con las multi-espectrales.

Asimismo, la banda térmica pasa de 120 a 60m, aunque para facilitar la superposición de capas dicha banda se vende re-muestreada a 30m.

Por otro lado, y de forma casi transparente para el usuario, se ha mejorado la calibración entre sensores –que ahora se hace en tiempo real a bordo del satélite y se facilita junto con las imágenes-, y se dispone de un mejor sistema de grabación. Con todo ello, afirman que antes de corregir la imagen, ya se consigue “una precisión geodésica de 400m a lo largo y ancho de la trayectoria orbital”.

Por otro lado, el Landsat 7, intentando adecuarse a los requerimientos de los estudios actuales, pretendió incorporar un nuevo sensor tratando de competir con las características de los exploradores de empuje de plataformas como SPOT o IRS. Este sensor denominado HRMSI (*High Resolution Multispectral Stereo Image*) presentaría una resolución de 5m con capacidad para cambiar el ángulo de observación a petición del usuario y reducir a 3 días la resolución temporal.

- Landsat 1, 2, 3 (1972, 75, 78): El 3 incluye una banda térmica.

- Landsat 4, 5: Se elimina la cámara RMV y se incluye TM para cartografía temática. Se tienen 7 bandas.



- Landsat 7 (1999): El Landsat 7 es el satélite operacional más reciente del programa Landsat, financiado por el gobierno de los Estados Unidos. El último satélite fue lanzado en abril de 1999 con un nuevo sensor denominado ETM+ (*Enhanced Thematic Mapper Plus*). Su operación es administrada por la NASA (*National Air and Space Administration*) y la producción y comercialización de imágenes depende de la USGS (*United States Geological Survey*).

CARACTERISTICAS de las BANDAS LANDSAT-7 ETM+				
Banda Numero	Rango Espectral (µm)	Líneas de Datos Escáner	de por Línea (bytes)	Bits por pixel
1	.450 - .515	16	6,600	8
2	.525 - .605	16	6,600	8
3	.630 - .690	16	6,600	8
4	.775 - .900	16	6,600	8
5	1.550 - 1.750	16	6,600	8
6	10.40 - 12.50	8	3,300	8
7	2.090 - 2.35	16	6,600	8
8	.520 - .900	32	13,200	8

TM y ETM+ Bandas Espectrales								
Anchura de banda (µm) Anchura Mínima – Anchura Máxima								
Sensor	Banda 1	Banda 2	Banda 3	Banda 4	Banda 5	Banda 6	Banda 7	Banda 8
TM	0.45 - 0.52	0.52 - 0.60	0.63 - 0.69	0.76 - 0.90	1.55 - 1.75	10.4 - 12.5	2.08 - 2.35	No existe
ETM+	0.45 - 0.52	0.53 - 0.61	0.63 - 0.69	0.78 - 0.90	1.55 - 1.75	10.4 - 12.5	2.09 - 2.35	.52 - .90
Región	Visible azul	Visible Verde	Visible Rojo	Infrarrojo Próximo	Infrarrojo Lejano	Térmico Lejano	Térmico próximo	Visible

Una imagen LANDSAT 7 ETM+ está compuesta por 8 bandas espectrales que pueden ser combinadas de distintas formas para obtener variadas composiciones de color u opciones de procesamiento. Entre las principales mejoras técnicas respecto de su antecesor, el satélite Landsat 5, en el Landsat 7 se destaca la adición de una banda espectral (Banda Pancromática) con resolución de 15 metros. También cuenta con mejoras en las



características geométricas y radiométricas y una mayor resolución espacial de la banda térmica para 60 m. Estos avances tecnológicos permiten calificar al LANDSAT 7 como el satélite más interesante para la generación de imágenes con aplicaciones directas hasta una escala de 1:25.000, principalmente en áreas rurales o territorios de grandes extensiones.

Banda Numero	Rango Espectral(μ)	Zona del Espectro	Resolución en el Terreno(m)
1	0.45 - 0.515 μ	Visible – azul	30
2	0.525 – 0.605 μ	Visible – verde	30
3	0.63 – 0.690 μ	Visible – rojo	30
4	0.75 – 0.90 μ	Infrarrojo Próximo (Visible)	30
5	1.55 – 1.75 μ	Infrarrojo Lejano	30
6	10.40 - 12.5 μ	Térmico Lejano	60
7	2.09 – 2.35 μ	Térmico Próximo	30
Pancromático	0.52 – 0.90 μ	Prácticamente todo el visible	15

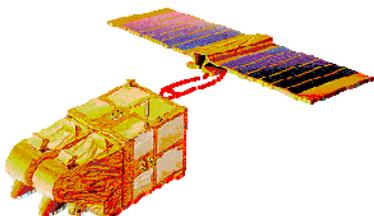
Las imágenes generadas por el Landsat 7 y adquiridas mediante el sensor ETM+ presentan una mejor relación costo-beneficio que los datos generados por satélites de resolución media (15 a 30 metros) actualmente ofrecidos en el mercado.

La Órbita del Landsat 7: El Landsat 7 puede adquirir imágenes en un área que se extiende desde los 81º de latitud norte hasta los 81º de latitud sur y, obviamente, en todas las longitudes del globo terrestre. Una órbita del Landsat 7 se realiza en aproximadamente 99 minutos, permitiendo al satélite dar 14 vueltas a la Tierra por día, y cubrir la totalidad del planeta en 16 días. La órbita es descendente, o sea de norte a sur; el satélite cruza la línea del Ecuador entre las 10:00 y 10:15 (hora local) en cada pasaje. El Landsat 7 está "helio-sincronizado", o sea que siempre pasa a la misma hora por un determinado lugar.



Un factor importante es que el período de revolución del LANDSAT 7 es igual al del Landsat 5 (16 días), y una imagen cubre igual área (185 x 185 km por escena). La conservación de estos parámetros técnicos facilita que el proceso de captura de imágenes se pueda realizar con la misma grilla de referencia (WRS2), lo que permite una perfecta integración entre el procesamiento de las imágenes del LANDSAT 7 con datos históricos del LANDSAT 5 existentes desde 1984. Esto es especialmente útil cuando es necesario utilizar los dos tipos de datos de un mismo lugar en forma simultánea, por ejemplo para un estudio multi-temporal.

SPOT:



El sistema SPOT es un programa espacial francés semejante al programa Landsat, concebido por el *Centre National d'Etudes Spatiales* (CNES), que lleva a bordo dos sensores de alta resolución (HRV - *HAUT Resolution Visible*). Estos sensores fueron concebidos para operar en dos modos diferentes. El modo multi-espectral permite la adquisición de datos en tres bandas del espectro electromagnético con una resolución espacial de 20 metros, y el modo pancromático con una banda de resolución espacial de 10 metros.

Una de las características principales de los instrumentos a bordo del SPOT es la posibilidad de observación "*off-nadir*". El sensor podrá ser direccionado para observar escenas laterales a la órbita en la que se encuentra el satélite en un momento determinado. Esta posibilidad de observación "*off-nadir*" aumenta los medios de obtener un aumento en el recubrimiento repetitivo de determinadas áreas. Otra ventaja de la vista "*off-nadir*" es la posibilidad de obtener pares estereoscópicos de determinadas áreas.



La luz proveniente de la escena alcanza un espejo plano que puede ser controlado a partir de las estaciones terrestres, variando en ángulos de +/- 0,6° hasta 27° con relación al eje vertical. La energía que alcanza el espejo plano es captada por una matriz lineal de detectores del tipo CCD (*Charge-Coupled Device*). Cada matriz consiste en 6000 detectores ordenados linealmente, formando lo que se denomina un "*push-broom scanner*" o sistema de barrido electrónico. Este sistema permite la toma instantánea de una línea completa en el terreno, perpendicularmente a la dirección del satélite en su órbita.

Las imágenes pancromáticas (blanco y negro) tienen una resolución de 10 x 10 metros, y se les usa comúnmente para cartografía, identificación, detección de cambios, y para actualización de mapas vectoriales. Las imágenes multi-espectrales (en el visible e infrarrojo) tienen una resolución de 20 x 20 metros, y son usadas para aplicaciones de uso y cartografía del terreno y para estudios de recursos naturales. Las ortoimágenes SPOT son ofrecidas en forma estándar con resoluciones de 7.5, 15 y 30 metros.

El ángulo de visión de cada sensor HRV (*High Resolution Visible*) puede ser ajustado para recoger datos hasta 27 grados a la derecha o a la izquierda del nadir. Esta habilidad de visión lateral permite que el mismo punto de la superficie terrestre pueda ser visto desde varias órbitas diferentes, permitiendo la adquisición de imágenes estereoscópicas.

Las imágenes SPOT contienen información de radiancia en la banda visible e infrarrojo, obtenida de sensores HRV (*High Resolution Visible*) montados en los satélites SPOT 1 y SPOT 2. Cada satélite dispone de dos sensores HRV con capacidad de registrar en dos modos: *Multispectral Linear Array* (MLA) o *Panchromatic Linear Array* (PLA). El primer modo captura datos en tres bandas: .050-0.59, 0.61-0.68, y 0.79-0.89 micrómetros. Las



tres bandas son registradas simultáneamente con una resolución en el terreno de 20 metros. En el segundo modo, las imágenes pertenecen a la banda de 0.51-0.73 micrómetros con una resolución de 10 metros en el terreno. Cada escena cubre normalmente 60 km².

NOAA



Los NOAA son satélites americanos de órbita polar. Están regidos y administrados por la *National Oceanic and Atmospheric Administration* y son los sucesores de los primeros satélites meteorológicos. Puede calificarse como el único sistema plenamente operativo para la adquisición y tratamiento histórico de la información meteorológica. Los satélites NOAA 14 y NOAA 15, lanzados respectivamente el 29 de mayo de 1994 y el 13 de mayo de 1998, vuelan a una altitud de 850 kilómetros y en una órbita inclinada con 99 grados respecto al plano ecuatorial. Cada 102 minutos hacen una órbita completa alrededor de la tierra, dando 14 órbitas por día. Estas órbitas son helio-síncronas. Los dos satélites están fuera de fase; se vuela por encima de una misma área por lo menos 4 veces al día, con un intervalo de aproximadamente 6 horas.

Los NOAA están equipados con un radiómetro (*Advanced Very High Resolution Radiometer*) con una muy alta resolución (1.1 kilómetros a la vertical del satélite), que permite captar la temperatura que emiten las superficies de los objetos en la Tierra. Explora una banda terrestre de 3000 kilómetros de ancho.



El sensor AVHRR del satélite NOAA provee de imágenes de 4 km de resolución (sistema APT) e imágenes de alta resolución de 1 Km (sistema HRPT). En un sistema receptor HRPT se pueden adquirir simultáneamente los cinco canales de datos de órbita polar para un área de interés. Estos datos incluyen los canales visibles, infrarrojo cercano (reflejado), y tres canales del infrarrojo termal (emitido). Desde hace casi cuatro décadas hasta nuestros días, permite un estudio prácticamente sistemático de la superficie terrestre.

Su principal objetivo es proporcionar información para la predicción del tiempo, y esto lo consigue con un ciclo orbital de tan solo 12 horas, para nuestras latitudes. Teniendo en cuenta la sincronización entre satélites redundantes (numerados como pares los principales, e impares los secundarios), este ciclo se reduce a la mitad. Es decir, cada 6 horas podemos disponer de imágenes.

Estas imágenes son de libre difusión, correspondiendo a un área de 3.000 km de lado, fruto de un ángulo de barrido de 55.4º a cada lado de la traza. Esta característica hace corresponder un IFOV de 1 km² para los píxeles en el nadir y 2.4 x 6.0 km a los píxeles extremos. Estos últimos adolecen de gran cantidad de ruido causado por el largo recorrido atmosférico que debe recorrer la radiación, por lo que en estudios normales se suele prescindir de un ángulo mayor de 30º.

Además de su gratuidad (<http://www.saa.noaa.gov>), que es un buen aliciente a la hora de abordar cualquier estudio, las imágenes NOAA presentan una resolución radiométrica de 210 en sus cinco bandas espectrales, configuradas de tal forma que son satisfactorias para los estudios para los que se implantó (estudio de nubes, línea de separación tierra-agua, extensión de superficies de nieve-hielo, índices de vegetación global, etc.).



Por otro lado, y cada vez con mayor profusión, se están acometiendo estudios no relacionados tan estrechamente con la predicción climatológica, que si bien no son estudios de detalle, sí permiten estudios globales con gran dinamismo (desertificación, deforestación, cobertura de suelos a nivel continental, inventario de cosechas, etc.).

Las imágenes se presentan en tres formatos diferentes: LAC (*Local Area Coverage*) o HRPT (*High Resolution Picture Transmission*) con resolución de 1km, GAC (*Global Area Coverage*), re-muestreo del anterior a 4 km, y GVI (*Global Vegetation Index*) que es un NDVI con 16 km de píxel.

Asimismo, los satélites NOAA disponen de otro sensor denominado TOVS (*TIROS Operational Vertical Sounder*), que consta de tres instrumentos distintos y que permite el estudio sistemático de perfiles verticales de temperatura en la atmósfera.

ERS 1



Bajo el acrónimo **ERS 1** (*European Remote Sensing Satellite*) se encuentra el primer satélite europeo de teledetección. Lanzado en 1991, contiene equipos de radar diseñados inicialmente para complementar las imágenes de sensores pasivos, sobre todo en océanos y criosfera.

El equipo más importante es un sensor activo de microondas denominado AMI (*Active Microwave Instrument*) que combina las funciones de un radar de apertura sintética (SAR) en la banda C, con una resolución espacial de 26 m en profundidad, y entre 6 y 30 m en acimut (nota: en cada referencia se suele dar una resolución espacial diferente para este satélite, que fluctúa entre 18-20-30m. Esta mezcla debe, sin duda, a la modificación del ángulo de incidencia en los sensores activos que hace variar la resolución espacial).



Las imágenes adquiridas por el SAR suministran datos de una franja de 100 x 100 km con una resolución espacial de 30 metros.

Una antena de 10 x 1 metros emite y recibe un haz de microondas en la región de 5,3 GHz (banda C), con polarización VV y un ángulo de incidencia de 23 grados.

El ángulo medio de incidencia para el ERS 1 es de 23°, con un área cubierta de 102km.

Admite lo que se conoce como modo 'olas', dirigido directamente al estudio del oleaje marino. Su resolución espacial es la misma, pero su área de adquisición se queda en 0.6 x 12km.

El propio AMI consta de un dispersómetro, que permite medir la dirección y velocidad del viento en el océano. Para ello utiliza tres antenas que emiten pulsos con distintos ángulos, con una resolución aproximada de 45 km y una cubierta total de unos 500 km.

Aparte del AMI, dispone de otro altímetro radar (RA) trabajando en la banda K. Envía un pulso vertical, y en función del tiempo que tarda en ser recogido de nuevo, es capaz de calcular la altura del oleaje con una precisión de 10 cm. Es posible su uso en otras coberturas de gran reflexión, como pueden ser campos de hielo.

El ATSR (*Along Track Scanning Radiometer*) es un equipo doble de barrido térmico con el fin de evaluar de forma precisa la temperatura del agua del mar y del techo nuboso, así como el vapor de agua atmosférico y la temperatura de la superficie terrestre. Cuenta con cuatro bandas en el MIR y TIR, obteniendo una precisión radiométrica entre 0.1-0.04K con resolución espacial de 1km en el nadir, en superficies de 500 x 500 km.

ERS 2

Este satélite europeo de teledetección fue lanzado el 21 de abril de 1995 por el Ariane 4 en una órbita polar circular a 780 km de altura.





Como añadidos, el ERS 2 dispone del GOME (*Global Ozone Monitoring Experiment*) para el control de la capa de ozono y demás componentes atmosféricos, e incorpora la nueva versión ATSR 2 con tres nuevas bandas en el visible y NIR.

Entre sus aplicaciones están las ya reseñadas, a las que se añade el hecho de incorporar las nuevas bandas en el visible: altitud del oleaje, dirección y velocidad del viento, bancos de hielo, recursos pesqueros, medidas de fitoplancton, contaminación química, etc. Y a todo ello se le han incorporado la obtención de MDE y el cálculo de subsidencias con precisiones centimétricas.

Sus instrumentos principales son: radar altimétrico (cartografía la superficie del mar); un difusómetro (mide la superficie marina así como la velocidad y la dirección del viento); un radar de apertura sintética (para imágenes de alta resolución de la superficie marina); instrumentos para la determinación de la concentración de ozono; sonda de microondas; y GOME (*Global Ozone Monitoring Experiment*).

Ambos satélites (ERS 1 y ERS 2) cuentan con instrumentos comunes: AMI (*Active Microwave Instrument*); ATSR (*Along-Track Scanning Radiometer and Microwave Sounder*); DEM (*Digital Elevation Models*).

3. Base de datos

En el sentido más amplio, se podría considerar que una base de datos es simplemente un conjunto de información. Existen numerosas definiciones de lo que es una base de datos y de lo que es un Sistema Gestor de Bases de Datos (S.G.B.D.). De hecho, a menudo se confunden ambos términos dada la relación biunívoca existente entre ambos. La definición más aceptada dice algo así como: «**En esencia, una base de datos no es más**



que una colección de información que existe a lo largo de un período de tiempo, a menudo de varios años. Más claramente, el término base de datos se refiere a una colección de datos gestionada por un Sistema Gestor de Bases de Datos, S.G.B.D., o simplemente Sistema de Bases de Datos». (Ullman: *A first course in database systems*).

Silberschatz extiende un poco más el concepto de base de datos, incluyendo en el término a los programas para acceder a estos datos, aunque no deja claro si son sólo los programas componentes del S.G.B.D., o incluye también a los programas de aplicación, hecho que otros autores como Julio Carmona dan por sentado, aunque con mucha menos rigurosidad. Nosotros consideraremos como base de datos a la colección de datos junto con el S.G.B.D.

Idealmente, el S.G.B.D. debe poseer una serie de características indispensables para satisfacer a los usuarios:

- Debe poseer un lenguaje de definición de datos que permita fácilmente la creación de nuevas bases de datos, así como la modificación de su estructura.
- Debe poseer un lenguaje de manipulación de datos que permita la inserción, eliminación, modificación y consulta de los datos de la base, de la forma más eficiente y conveniente posible.
- Debe permitir el almacenamiento de enormes cantidades de datos (miles de millones de caracteres) sin que el usuario perciba una degradación en cuanto al rendimiento global del sistema.
- Debe permitir la gestión segura de los datos, con respecto a accesos no autorizados y a accidentes producidos por los dispositivos mecánicos o electrónicos que soportan los datos almacenados.



- Debe permitir el acceso simultáneo por parte de varios usuarios, impidiendo además que dichos accesos concurrentes den lugar a datos corruptos o inconsistentes.
- Debe suministrar la independencia física de los datos que asegure que, sea cual sea la estructura de los datos en los dispositivos electromecánicos de almacenamiento, el usuario y las aplicaciones los percibirán siempre de manera uniforme y útil.

Estas características dan lugar a otra definición de base de datos, quizás más orientada a la moda actual en informática: el diseño orientado a objetos, que consiste en hacer programas viendo el mundo como un conjunto de objetos que se relacionan entre sí para conseguir un objetivo común.

Según esta visión, podemos considerar una base de datos como **«una organización coherente de datos permanentes, y accesibles para usuarios concurrentes»**. Esta definición lleva implícito el concepto de S.G.B.D., ya que es éste quien se debe encargar de almacenar los datos (y por tanto de hacerlos permanentes y no volátiles), de gestionar su integridad (y por tanto de que sean coherentes), y de controlar a los múltiples usuarios (permitiendo así la concurrencia).

Por ejemplo, una base de datos muy sencilla podría ser una agenda de direcciones en la que anotemos datos de nuestros amigos. Tendríamos una ficha para cada uno de estos amigos. En cada ficha, a su vez, existirá una serie de apartados, como el nombre, la dirección, el teléfono, etc.

Esto nos permite introducir dos primeros conceptos que utilizaremos: cada una de estas “fichas” recibe el nombre de “registro”, y cada uno de los “apartados” que componen las fichas se llama “campo”.



El conjunto de las fichas (registros) forma un “bloque” de información, que llamaremos “**tabla**”, y que se suele representar escribiendo cada ficha (registro) en una fila y cada apartado (campo) en una columna.

Pero en la práctica, una “base de datos” real suele estar formada por más de una tabla. Por ejemplo, la base de datos que utiliza una empresa “normal” para su gestión deberá almacenar datos sobre clientes, proveedores, artículos, facturas, etc.

Cada uno de estos “bloques” de datos será una tabla, y estas tablas estarán relacionadas entre sí (por ejemplo: un artículo será suministrado por un cierto proveedor, y ese artículo aparecerá en ciertas facturas, cada una de las cuales corresponderá a un cierto cliente).

Todo este conjunto de información que forman las tablas y las relaciones entre ellas (y alguna cosa más que ya veremos) será nuestra “base de datos”. En concreto, se tratará de lo que se conoce como una “base de datos relacional”.

En resumen, podemos decir que una base de datos es un programa que se encarga de gestionar todo el tratamiento de entrada, salida, protección y elaboración de la información que almacena. Aunque aparentemente podamos pensar que una base de datos son ficheros donde se almacena información, en realidad esto no es así.

El corazón de una base de datos es el motor, que es el programa que debe estar ejecutándose en una máquina para gestionar los datos. Además de este programa y de los archivos con datos, existen otras utilidades auxiliares, como programas para realizar copias de seguridad, intérpretes SQL, etc.

Se pueden clasificar las bases de datos de dos formas:



Desde el punto de vista de la organización lógica:

- **Jerárquicas.** (Progress)
- **Relacionales.** (Oracle, Access, Sybase...)

Desde el punto de vista de la arquitectura y el número de usuarios:

- **De escritorio** (dBase, Access, Paradox...)
- **Cliente/servidor** (Oracle, Sybase...)

Funciones de las bases de datos

- Permitir la introducción de datos por parte de los usuarios (o programadores).
- Salida de datos.
- Almacenamiento de datos.
- Protección de datos (seguridad e integridad).
- Elaboración de datos.

Componentes de un sistema de base de datos.

Hasta ahora hemos visto una idea general de lo que es una base de datos, y la necesidad de que exista un gestor central que separe los datos en sí de los usuarios, y que facilite una serie de tareas, así como que se asegure de que se cumplen determinados requisitos indispensables para el buen funcionamiento del sistema.

Ahora vamos a desmenuzar un poco más el concepto de base de datos, y vamos describir sus componentes básicos, justificando además la necesidad de cada uno de ellos.

Datos.

Efectivamente, una base de datos no tiene sentido si no está compuesta por datos. Lo que no está tan claro es la forma en que estos datos se deben disponer, qué datos se deben almacenar, y cómo los debe entender la máquina.



La disposición de los datos depende del ámbito de aplicación concreto en que se enmarque la base de datos. No es lo mismo una base de datos que almacene un dibujo vectorial de monumentos históricos que una base de datos para almacenar las reservas de clientes en un hotel. Lo que varía fundamentalmente en uno y otro caso es la forma en que los datos se relacionan entre sí, y el tipo de accesos a la información que va a realizar el usuario.

Además, los datos deben disponerse de manera que las consultas sean lo más eficientes posible, evitando a la vez la existencia de datos duplicados que pueden dar al traste con la coherencia de la base de datos.

De esta forma, no sólo se consideran datos aquellos que el usuario desea almacenar, sino toda estructura de apoyo que el sistema necesite para hacer más eficiente una consulta. Por ejemplo, supongamos que trabajamos en un hotel cuya base de datos posee un fichero con todas las agencias de viajes que contratan nuestros servicios de alojamiento en alguno de sus paquetes de viajes.

Supongamos que para facilitar los accesos más comunes tenemos las fichas de tales agencias ordenadas por nombre. ¿Qué ocurre si queremos acceder a todas las agencias de Egipto para mandarles una carta felicitándolas por el año nuevo musulmán?

Resulta inaceptable ir una por una a todas las agencias de nuestro fichero, ya que esto nos llevaría un tiempo desorbitado. La solución informática a este problema consiste en crear un fichero aparte, cuyas fichas contienen sólo dos campos: País de la agencia, y Nombre de la agencia. Las fichas de este fichero estarán ordenadas por País, y dentro de cada país, por orden alfabético por el Nombre de la agencia.

Para acceder a las agencias de Egipto, no accedemos al fichero principal, sino que nos dirigimos a este fichero auxiliar ordenado por país, y de ahí podemos averiguar



directamente cuáles son los nombres de las agencias de Egipto. A continuación nos vamos al fichero principal, accediendo sólo a las fichas cuyos nombres hemos extraído anteriormente del auxiliar, y de estas extraemos la dirección, lo que permite, finalmente, el envío de las cartas. El fichero auxiliar es lo que se llama un índice del fichero principal.

La gran ventaja de este método es que el mantenimiento del fichero auxiliar, cuando trabajamos con un ordenador, lo realiza automáticamente el sistema, lo que nos ahorra la tarea de tener que escribir explícitamente una ficha auxiliar cada vez que se inserta una nueva agencia en el fichero principal. De esta forma se obtiene una gran eficiencia en las consultas, sin que el usuario tenga que trabajar más: el ordenador trabaja por nosotros.

Pues bien, este fichero aparte, creado con el único objetivo de facilitar el acceso a la información, también lo consideramos datos, aunque deban ser gestionados automáticamente, e incluso algunos usuarios no tengan ni la menor idea de su existencia.

Por otro lado, es muy importante decidir qué datos son los que se van a almacenar. A menudo ocurre que en una base de datos se almacenan las facturas con una antigüedad inferior a x años. El número de años que se desea almacenar una factura electrónicamente puede ser una decisión crucial dependiendo de la capacidad de almacenamiento de nuestra máquina y del volumen de facturas que se expidan anualmente.

Además, si casi no se necesita acceder a dicha información histórica, puede ser más interesante destruir electrónicamente esa información, y almacenarla exclusivamente en papel, con lo que el ahorro de espacio en disco puede contrarrestar el esfuerzo de hacer las consultas manualmente.



Contrariamente, cuando se crea la base de datos, pueden parecer innecesarios ciertos datos de un cliente, como por ejemplo su fecha de nacimiento. Posteriormente, si trabajamos en una agencia de viajes y creamos un paquete de viajes de aventura por la selva del Amazonas, puede interesarnos promocionarlo sólo a aquellos clientes menores de 35 años, con lo que nos damos cuenta de que habría sido buena idea almacenar la fecha de nacimiento. Aunque el S.G.B.D. nos permite modificar la estructura de la base de datos para almacenar este nuevo campo, introducirlo para todos los clientes es una tarea ardua y fatigosa.

Por último, también es interesante, por cuestiones de seguridad y aprovechamiento de los recursos del ordenador, decidir en qué formato se van a almacenar los datos: si se van a almacenar encriptados para que nadie los pueda copiar, o si se van a codificar o a comprimir de alguna forma para hacer que ocupen menos espacio.

Existe toda una teoría sobre compresión y codificación de datos que puede hacer que, dependiendo de las características de los datos a almacenar, se requiera mucha menos memoria de la necesaria siguiendo los métodos convencionales de almacenamiento.

Metadatos.

La evolución de las bases de datos ha hecho que los informáticos se den cuenta de determinados problemas, y que los hayan ido subsanando a medida que perfeccionaban sus productos.

Uno de los problemas que se producían se relaciona con la propia evolución de la base de datos. Desde el momento en que se crea una base de datos, hasta el momento en que se desecha porque se compra un sistema mejor o se instala una nueva base de datos, la estructura de la base de datos (o sea, los datos que se



deciden almacenar y la estructura con que se almacenan) cambia a medida que cambian las necesidades sobre la información a obtener de la base de datos, de manera que puede ser necesario introducir nuevas fichas, nuevos campos en fichas ya existentes, eliminar campos, modificar lo que se almacena en determinados campos, e incluso reestructurar la base de datos entera, crear ficheros de apoyo, etc.

Dado que la base de datos que solucione unas necesidades concretas puede adoptar muchas formas posibles, es muy interesante poseer algún lugar que indique al personal encargado de mantener la base de datos cuál es el objetivo de cada dato particular almacenado en la base, así como en qué aplicaciones es utilizado, con qué propósito, si es un dato fundamental, o si puede ser omitido por el que introduce los datos, etc.

De esta forma, antes de modificar el esquema o estructura de la base de datos, el departamento de procesamiento de datos debe consultar esta información sobre los datos de la base con cuidado para no cometer errores graves que repercutan sobre el buen funcionamiento de todo el sistema.

Esta información que el sistema guarda sobre los datos almacenados, es lo que se llaman metadatos. Es más, estos metadatos se almacenan como otra base de datos propiamente dicha, y puede ser gestionada y consultada como tal. Estos metadatos suelen conformar lo que se da en llamar diccionario de datos.

El sistema gestor de bases de datos.

Según Silberschatz, podemos considerar un S.G.B.D. como **«un programa que proporciona la interfaz entre los datos de bajo nivel almacenados en la base de datos, y los programas de aplicación y consultas hechos al sistema»**.



Ya hemos indicado las características propias de una base de datos, la mayoría de las cuales recaen precisamente sobre el S.G.B.D. Para dejarlas más patentes, las reunimos y ampliamos aquí:

- a) **Interactuar con el Sistema Operativo.** Como se ha indicado, el S.G.B.D. no es más que un programa. El Sistema Operativo es el programa principal que se encarga de controlar que el ordenador funcione bien, entre otras cosas permitiendo el acceso a los dispositivos de entrada y de salida como el teclado, el ratón, el monitor, y los dispositivos de almacenamiento: el disco duro, las disqueteras, el CD-ROM, las cintas magnéticas, etc.

Así, para asegurar que no pasen cosas raras, el único que puede tocar estos dispositivos es el Sistema Operativo (S.O.). Dado que el S.G.B.D. necesita almacenar datos en el disco duro, debe interactuar con el S.O. para poder acceder al disco duro y almacenar allí los datos que se quieren. Igualmente, debe ponerse en contacto con el S.O. siempre que desee recuperar algún dato de estos dispositivos.

- b) **Mantener la integridad.** Como ya se ha dicho, debe mantener las restricciones de integridad propias de la aplicación concreta que sea. Por ejemplo, evitar que la edad de un cliente supere los 90 años.
- c) **Mantener la seguridad.** Evitar accesos fraudulentos a los datos, así como la extracción de información codificada-, y permitir las copias de seguridad. Dado que un ordenador no es un sistema infalible, y puede romperse por causas propias (fallo de un circuito) o ajenas (aumento de la tensión en la red eléctrica), es posible que los datos almacenados por él lleguen a corromperse con la



consiguiente pérdida de información y los problemas que ello puede acarrear a la empresa.

Para evitar estos desagradables resultados, es buena idea efectuar una copia de los datos a un dispositivo auxiliar de almacenamiento, pensado precisamente para guardar fiel copia del contenido de la base de datos en un momento determinado. Si los datos originales se destruyen, bastará volcar la copia sobre el disco duro del ordenador central, con lo que los datos volverán a tomar la misma forma que cuando se efectuó la copia. De esta manera, para que la base de datos recupere la forma que tenía en el momento en que quedó destruida, bastará con efectuar los cambios que se hicieron en el tiempo transcurrido desde la copia de seguridad que se acaba de volcar.

- d) **Controlar la concurrencia.** Como ya se explicó anteriormente, debe permitirse el acceso simultáneo a los datos por parte de varios usuarios, lo que conlleva numerosos problemas de coherencia y coordinación. El S.G.B.D. debe controlar que la información representada por los datos al final de cada acceso de usuario siga siendo consistente.
- e) **Suministrar mecanismos que faciliten la interacción con la base de datos.** Estos mecanismos suelen venir dados en forma de lenguajes de manipulación y definición de datos. Además, suministran independencia de los datos, en el sentido de que, a pesar de la evolución del esquema de los datos, las aplicaciones deben sufrir las mínimas modificaciones imprescindibles. Por ejemplo, si las aplicaciones antiguas están pensadas para trabajar sobre números de teléfono expresados mediante dígitos, ¿qué ocurre si se decide cambiar todos los números a formato textual? En este caso, debe haber un mecanismo que oculte a las aplicaciones antiguas el nuevo formato de los números de teléfono,



y los haga ver en el formato antiguo; en definitiva, debe haber algo que suministre a las aplicaciones antiguas una visión ligeramente distinta de lo que hay realmente almacenado en la base de datos.

Componentes de un S.G.B.D.

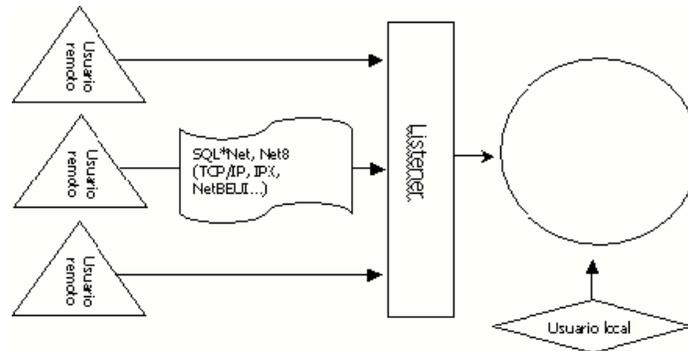
Una base de datos consta de varios componentes:

- a) **Motor:** el programa ejecutable que debe estar en memoria para manejar la base de datos. Cuando este programa está ejecutándose se dice que la base de datos está levantada (*startup*); en caso contrario se dice que la base de datos está bajada (*shutdown*).

- b) **Servicio de red:** es un programa que se encarga de establecer las conexiones y transmitir datos entre cliente y servidor, o entre servidores.

- c) **Listener (Escuchador):** es un programa residente en memoria que se encarga de recibir las llamadas que llegan a la base de datos desde la red, y de pasárselas a esta. Una base de datos que no tenga un *listener* cargado no podrá recibir llamadas remotas. El *listener* se comunica con el servicio de red.

- d) **Utilidades:** Programas de utilidad como pueden ser:
 - i. Intérpretes de consultas.
 - ii. Programas de administración de bases de datos.
 - iii. Programas de copia de seguridad.
 - iv. Monitores de rendimiento.



A todo el conjunto de la base de datos se le denomina RDBMS: **Relational DataBase Manager System**, es decir, **Sistema de Gestión de Bases de Datos Relacionales**.

El primer fabricante en diseñar un RDBMS fue IBM, pero fue Oracle, en 1979, la primera empresa en hacer una implementación comercial de un sistema de bases de datos relacionales.

Usuarios de la base de datos.

Hasta ahora hemos entendido como usuarios a aquellos individuos que esperan obtener algún resultado de la base de datos. Sin embargo, también podemos considerar usuario a toda persona encargada del mantenimiento y puesta a punto de los datos. De esta forma, podemos dividir a los usuarios en dos grandes grupos:

- a) **Administrador.** Es aquella persona o conjunto de personas encargada (exclusivamente, en la mayoría de los casos) del control global de la base de datos.

Sus tareas son:



- **Definición del esquema:** Como se ha indicado, el esquema de una base de datos está compuesto por el tipo de fichas que almacena, el formato que toman cada una de ellas, así como sus características y reglas para ser rellenas.

El administrador es, pues, quien decide qué se va a almacenar, y cómo. La información del esquema debe ser almacenada como metadatos, o sea, información sobre el formato de los datos almacenados.

- **Definición de la estructura de almacenamiento y de los métodos de acceso.** Se indicó anteriormente que el S.G.B.D. interactuaba con el S.O. para poder acceder a los datos guardados sobre los dispositivos de almacenamiento, o sea en los ficheros. Pues bien, estos ficheros están a su vez estructurados con base en unidades más pequeñas con el objetivo de facilitar el acceso a los datos por parte del S.O. Es tarea del administrador decidir las características de estas unidades mínimas de almacenamiento a fin de agilizar la recuperación de los datos, y de mejorar los accesos en función de las características propias de las consultas más frecuentes por parte de los usuarios.

También con base en las consultas más frecuentes, el administrador debe decidir sobre la necesidad de mantener estructuras de apoyo que hagan más fácil la recuperación de determinados datos. Cuando hablamos de estructuras de apoyo, nos referimos a los índices que veíamos en un punto anterior. Creando los índices apropiados, la velocidad de determinadas consultas puede aumentar vertiginosamente, incrementando así la satisfacción de los usuarios en el manejo de la base de datos.

Por otro lado, veíamos que los índices son mantenidos automáticamente por el sistema, lo que puede producir ralentizaciones ante algunas operaciones de inserción, eliminación o modificación de datos, sin que aparentemente haya



motivo alguno. Además, un índice ocupa un espacio cuyo tamaño depende del fichero original, por lo que la creación de demasiados índices puede requerir una cantidad de memoria en nada proporcional a los beneficios obtenidos de su utilización.

Modificación del esquema y de la organización física. A medida que la empresa que utiliza la base de datos evoluciona para adecuarse a las condiciones cambiantes del mercado, es necesario adecuar dicha base de datos a las necesidades cambiantes que se producen. De esta forma, debe ser posible alterar el esquema de los datos, así como la forma de almacenamiento para hacer que las nuevas necesidades sean satisfechas de forma eficiente.

Asimismo, es interesante que las aplicaciones antiguas sigan funcionando con un mínimo de modificaciones, por lo que es de desear que, a pesar de dichas modificaciones, las aplicaciones antiguas tengan una visión alterada de los datos, pero adecuada a su forma de funcionamiento. Por supuesto, estas visiones o vistas deben ser coherentes con los datos almacenados en el esquema nuevo, y no tergiversar el contenido de los datos.

- **Concesión de permisos y privilegios para el acceso a los datos.** Asignación de prioridades a los datos. Para gestionar la seguridad en la base de datos, hemos comentado que es necesario que los datos dispongan de una serie de niveles de importancia, de manera que sólo puedan ser accedidos por los usuarios con un privilegio superior o igual a dicho nivel. Pues bien, es el administrador quien se encarga de asignar dichos niveles y privilegios, en función de las características particulares de cada usuario. Asimismo, debe asignar inicialmente las claves de acceso a los usuarios, ya que cada uno de ellos sólo puede acceder al sistema mediante una contraseña particular.



- **Especificación de restricciones de integridad.** Los datos almacenados en la base de datos guardan una relación entre sí en función de las necesidades que pretenden cubrir, de manera que la información que se pueda extraer de ellos sea útil y coherente. Esta coherencia viene determinada por un conjunto de reglas o restricciones que los datos deben poseer; por ejemplo, ningún cliente debe tener pendientes de pago facturas cuyo monto supere el crédito máximo permitido. El administrador debe indicar estas restricciones de integridad, así como modificarlas o flexibilizarlas a medida que evolucionan las necesidades de la empresa. A menudo, estas restricciones se suelen considerar incluso como parte integrante del esquema, con lo que también quedan reflejadas en el diccionario de datos.

- b) **Otros usuarios.** El resto del personal que interactúa con la base de datos puede considerarse simplemente como usuarios, aunque también podemos distinguir distintos tipos entre ellos:
 - 1) **Programadores de aplicaciones.** Suelen formar parte del departamento de procesamiento de datos. Reciben peticiones de otros usuarios en forma de necesidades de acceso a los datos, y se encargan de escribir los programas que satisfacen dichas necesidades. Normalmente estos programas están escritos en un lenguaje de programación convencional (Pascal, Basic, C, etc.) en el que se insertan órdenes especiales que el S.G.B.D. es capaz de comprender. De esta forma, el S.G.B.D. suministra los datos, y el lenguaje convencional (también llamado anfitrión porque alberga las sentencias reconocidas por el S.G.B.D.) los procesa, presenta al usuario, modifica, etc.



- 2) **Usuarios directos.** Son aquellos que interactúan con la base de datos haciendo uso directamente del lenguaje que proporciona el S.G.B.D. En capítulos posteriores estudiaremos el SQL como principal lenguaje de manipulación y definición de datos. El administrador debe tener cuidado con la prioridad que asigna a estos usuarios, ya que al tocar directamente la base de datos, pueden alterar su estructura si disponen de contraseñas de alta prioridad, o si por error se les ha suministrado una prioridad demasiado grande.

- 3) **Usuarios indirectos.** Son aquellos que no saben nada de la base de datos, excepto que existe. Interactúan con ella a través de las aplicaciones desarrolladas por los programadores, y son incapaces de acceder a los datos directamente a través del lenguaje del S.G.B.D. El administrador debe prestar especial atención a que este tipo de usuarios no acceda directamente a los datos por los problemas que su inexperiencia pueda acarrear.

Elementos de seguridad.

El administrador debe conocer en profundidad los elementos de seguridad que suministra el S.G.B.D. de que dispone, y sacar el máximo partido posible de ellos. En general, se pueden tener niveles de acceso clasificados por:

- **Las fichas o tuplas a que se tiene acceso.** En una empresa grande y ampliamente informatizada, cada usuario debe poder acceder exclusivamente a los datos que competen a su tarea. El departamento de contabilidad no tiene porqué acceder a la información sobre reservas de alojamiento, aunque sí a las peticiones de tintorería, para hacer los cargos correspondientes a cada cliente. A su vez, el recepcionista



solamente accederá a las reservas efectuadas, y podrá abrir una factura de alojamiento en la que se irán insertando por parte de cada sección del hotel los cargos de los servicios prestados al cliente, pero no podrá tener en ningún momento acceso a las nóminas, ya que ello es responsabilidad de la habilitación.

De esta forma vemos que cada usuario sólo debe tener acceso a un tipo de información, o lo que es lo mismo, a un tipo de fichas determinado. Por tanto, debe existir un mecanismo de seguridad que restrinja el ámbito de acceso de cada usuario en función de sus competencias.

- **Las operaciones que se pueden realizar sobre las fichas o tuplas.** Siguiendo con el caso anterior, la sección de contabilidad tendrá acceso a las facturas a clientes, pero no podrá modificarlas, sino tan sólo consultarlas a efectos contables para generar el balance de cada mes. Por otro lado, las diferentes secciones de cafetería, restaurante, tintorería, sauna, joyería, etc., sí deben poder acceder a las facturas de los clientes para realizar en ella los cargos pertinentes. Sin embargo, no podrán crear facturas nuevas, si sólo pueden atender a clientes que se alojen en el hotel.

Por tanto, no sólo es importante el acceder o no a los datos, sino también la forma en que este acceso se produce, en función de las características propias de la sección: la habilitación sólo puede consultar facturas de clientes; la recepción puede crear y modificar, así como cerrar una factura; la tintorería sólo puede añadir cargos; la joyería puede crear y añadir cargos, etc.

En general, las operaciones que se pueden efectuar sobre un tipo de fichas se agrupan en cuatro grandes bloques: Altas, Bajas, Modificaciones y Consultas.

- **El acceso al diccionario de datos y a la estructura de la base de datos.** Como se ha comentado anteriormente, los metadatos almacenados en el diccionario de datos, que almacenan información sobre la estructura de la base de datos, son gestionados,



a su vez, como si se tratase de una base de datos especial. Sin embargo, dada su primordial importancia, su acceso debe estar muy restringido, ya que cualquier modificación puede dar lugar a resultados desastrosos en la base de datos: pérdida de información, corrupción en los datos, falta de integridad, etc. Por ello, es necesaria la existencia de prioridades o privilegios especiales que sólo permitan el acceso al personal que compone la administración de la base de datos, que es el único capacitado para modificar estos metadatos.

Lenguajes de bases de datos.

La interacción del usuario con la base de datos debe efectuarse a través de alguna técnica que haga fácil la comunicación, y que permita al usuario centrarse en el problema que desea solucionar, más que en la forma de expresarlo con las técnicas que se le suministran. La mejor forma de alcanzar este objetivo es darle un lenguaje parecido al lenguaje natural, que le permita expresar de forma sencilla los requerimientos. En función de estos requerimientos, podemos tener fundamentalmente dos tipos de lenguajes para comunicarnos con el S.G.B.D.

Lenguaje de definición de datos (DDL). Un esquema de BD se especifica por medio de un conjunto de definiciones que se expresan mediante un lenguaje especial llamado lenguaje de definición de datos (*Data Definition Language, DDL*). La estructura de almacenamiento y los métodos de acceso se especifican por medio de un conjunto de definiciones en un tipo especial de DDL llamado lenguaje de almacenamiento y definición de datos (**diccionario de datos**). El resultado de la compilación de estas definiciones es un conjunto de instrucciones que especifican los detalles de implementación de los esquemas, que normalmente se esconden a los usuarios.



Este lenguaje es utilizado en exclusiva por el administrador de la base de datos, ya que permite la construcción de frases o sentencias que le indican al S.G.B.D. las características particulares de la base de datos sobre la que se está trabajando, así como la creación de nuevas bases de datos. La creación de esquemas y su modificación, la creación y supresión de índices, la especificación de unidades de almacenamiento en los ficheros, así como la asignación y privación de prioridades, se realizan a través de este lenguaje. Además, permite la creación y recuperación de copias de seguridad, así como la importación de datos desde sistemas antiguos, y la exportación a nuevos sistemas que vayan a sustituir al actual.

Lenguaje de manipulación de datos (DML). El lenguaje de manipulación de datos es el que usan los usuarios directos para efectuar sus operaciones sobre la base de datos. Como se indicó, estas operaciones son básicamente de **inserción, eliminación, modificación y consulta de datos**, aunque también se pueden introducir capacidades para crear visiones de los datos que faciliten otros accesos.

Los usuarios directos interactúan con el S.G.B.D. a través de este lenguaje, mediante una interfaz agradable y fácil de usar. Los programadores de aplicaciones emplean el **DML** dentro de un lenguaje de programación que les da potencia expresiva, es decir que les permite sacar cosas por pantalla de una forma agradable, facilitar las operaciones de los usuarios indirectos, etc.



Un lenguaje de manipulación de datos (*Data Manipulation Language, DML*) es un lenguaje que permite a los usuarios acceder o manipular los datos. Existen básicamente dos tipos:

- Procedimentales:** requieren que el usuario especifique qué datos se necesitan y cómo conseguirlos.
- No procedimentales:** el usuario debe especificar qué datos se necesitan, pero no cómo obtenerlos.

Los **DML** no procedimentales normalmente son más sencillos de aprender y usar; sin embargo pueden generar códigos que no sean tan eficientes como el producido por los lenguajes procedimentales. Esta dificultad puede remediarse a través de varias técnicas de optimización.

4. Sistema de Información Geográfica SIG

Historia

La contribución del pasado

Muchos de los algoritmos que actualmente permiten llevar el paisaje al lenguaje computacional provienen de una matemática que se remonta aproximadamente al año 1680. De estas escuelas matemáticas surgió la **topología**, ciencia matemática que permite estudiar las figuras y sus relaciones entre sí. Del mismo modo surgieron los cuatro principios de la naturaleza de los datos geográficos: “todo elemento geográfico tiene posición absoluta, posición relativa, figura geométrica y atributos”.



Los primeros SIG

Hacia la década de 1870 se organizó un sistema de información geográfica por parte de una empresa de trenes en Irlanda, que empleó la superposición de acetatos.

Hacia los años 50 del siglo XX aparecieron los primeros software de cartografía automatizada y las primeras bases de datos para manejar atributos en el computador. Luego, a finales de los años 60 surgieron sistemas que permitían integrar las bases de datos con las figuras, y esta facilidad fue puesta en práctica desde entonces.

El primer SIG que logró cierta eficiencia fue el SIG Canadá, en 1962, el cual fue orientado al manejo de bosques (administración de sus recursos naturales) y estaba estructurado sobre todo en polígonos.

Evolución de técnicas

Desde los años 60 y hasta mediados de los 70 se manejó un **modelo orientado a registros**: cada figura tenía un registro correspondiente pero no se podía establecer una relación entre las figuras.

A comienzos de los años 70 se desarrollaron algoritmos que permitían generar las posiciones relativas mediante topología en capas o *layers*. Esta técnica llamada **modelo orientado a capas** (o también orientado a BD) se impuso durante los años 80 y aún perdura en muchos estudios pues es la técnica más práctica y comercialmente distribuida.

En 1985 los ingleses crearon el **modelo orientado a objetos**, donde se considera el paisaje tal como lo es realmente: todo se conforma de partes, y las partes se integran y forman objetos.



Concepto de SIG

Existen diferentes definiciones para un sistema de información geográfica; una definición sintetizada sería: **“Herramienta informática para la manipulación y análisis de datos georeferenciados, orientada a la toma de decisiones”**.

Pero más detalladamente, podemos decir que un sistema de información geográfica es **un sistema de hardware, software y procedimientos diseñado para realizar la captura, almacenamiento, manipulación, análisis, modelado y representación de datos referenciados espacialmente para la resolución de problemas complejos de planificación y gestión.**

Componentes

Recurso técnico

Subsistema de entrada: Realiza la captura y transformación de datos análogos tales como mapas impresos, registros alfanuméricos en papel y observaciones de campo. Del mismo modo, convierte la información digital, proveniente de sensores remotos u otros sistemas de información, a una plataforma compatible con lenguaje computacional del SIG. Entre los dispositivos de entrada figuran: Tableros digitalizadores, escáneres o barredores, lectores magnéticos y láser, teclados, terminales y puertos, e Internet.

Subsistema de manejo: Es el subsistema que permite el almacenamiento, la ordenación y la recuperación de datos. Esta organización es posible gracias a programas conocidos como Sistemas Manejadores de Bases de Datos (SMBD) que permiten manejar datos espaciales digitales. Mediante las bases de datos y los SMBD se obtiene una administración de datos que permiten su consulta, tratamiento de datos derivados y su

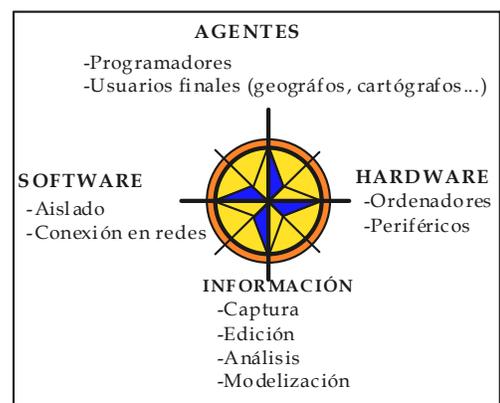


retroalimentación. Entre los dispositivos de almacenamiento figuran: discos duros, cintas magnéticas y unidades de compresión y *zip-back up* (archivos zip de respaldo).

Subsistema de análisis: Existen muchos análisis en SIG, que van desde la sencillez de la comparación de objetos según sus atributos, hasta complejos análisis de rutas eficientes en tiempo y distancia. Son análisis típicos en los paquetes de SIG el análisis espacial, el análisis de proximidad, el análisis de redes y el análisis en tercera dimensión, entre otros. El éxito de estas operaciones recae en la calidad y preparación de la información a ser analizada. Se requiere de una correcta conceptualización de las tareas de análisis previa a su ejecución.

Subsistema de salida: Es el subsistema que comprende la presentación de los datos y el despliegue de los resultados derivados del subsistema de análisis. La salida de datos corresponde tanto a despliegues gráficos (mapas, gráficas) como alfanuméricos (tablas, reportes). A su vez la salida puede generarse tanto en formatos análogos como digitales que puedan ser exportados mediante diversos medios a otro SIG u otro software similar. Entre los dispositivos de salida en SIG figuran: terminales y puertos de salida, impresoras, plotters, cintas magnéticas, discos de almacenamiento, medios ópticos.

- **Recurso humano:** Un SIG requiere un equipo humano cuya preparación no debe limitarse al conocimiento de los SIG mismos, sino que debe cubrir razonablemente las diferentes áreas implicadas en los análisis y campos de aplicación. El recurso humano lo comprenden tanto las personas capaces de conceptualizar y manejar las utilidades de la tecnología SIG



Severino Escolano



como también aquellas que actúan solamente en calidad de cliente. A su vez, los usuarios pueden ser internos o externos a un SIG específico.

- **Organización:** La implantación de un SIG puede desarrollarse exitosamente bajo el amparo de una voluntad institucional fuerte, decidida y convencida a fondo de las implicaciones de adoptar esta tecnología.
- **Datos:** Un SIG opera con datos geo-espaciales. Un dato geo-espacial es aquel que hace referencia a un espacio geográfico cuya ubicación se conoce (bajo un sistema de coordenadas). Los SIG almacenan la **localización** del dato, su relación espacial con otros datos (**topología**) y una descripción a través de sus **atributos** propios. Un SIG sin datos no es SIG, sino simplemente un software vacío.

Por otra parte, también existen diferencias entre las aplicaciones desarrolladas con SIG, en cuanto a su tamaño, objetivos y otros aspectos estructurales, como se resume en el siguiente cuadro.



Sistemas	Orientación	Tamaño	Tiempo	Características	Desarrollo
Grandes sistemas	-Propósito general (Catastro, sistemas de información territorial, sistemas de información urbana, etc.)	-Grande	-Largo plazo	-Planificación exhaustiva -Formatos -Procedimientos normalizados	- Instituciones -Grandes empresas
Sistemas medianos y pequeños	-Proyectos (gestión y planificación del comercio, de una explotación agraria, un servicio, etc.)	- Mediano - Pequeño	- Corto medio y plazo	-Planificación flexible	-Equipos -Particulares

Ventajas

A continuación enumeramos algunas de las ventajas que aportan los sistemas SIG:

Los SIG brindan el salto del mapa impreso en papel al manejo de mapas digitales, y el salto a la superposición digital.

A diferencia de la cartografía digital, que no va más allá de la ubicación de los objetos, los SIG no sólo nos permiten manipular los elementos de un mapa sino relacionar cada objeto con una información más amplia y establecer relaciones espaciales y de carácter.



Los SIG permiten análisis matemático y salidas gráficas para visualizar resultados parciales y finales de un trabajo.

Como los SIG manejan la base de datos por un lado y la presentación por otro, se pueden generar muchos mapas desde los mismos datos.

La naturaleza interdisciplinaria que orienta los trabajos en SIG se hace más fácil pues existe una conexión entre la información temática elaborada a priori por distintos especialistas y el manejo de un área de estudio.

Representación de la Información Geográfica. (IG)

Veremos en primer lugar los marcos de medida de los Sistemas de Información Geográfica (SIG) donde detallaremos conceptos básicos que nos servirán para comprender mejor los modelos que presentaremos en la sección de modelos de estructuras de datos, y su posterior almacenamiento en las máquinas que van a dar soporte a los SIG.

Definición de IG.

Los SIG trabajan con información geográfica (IG); el verdadero valor de los SIG estriba en su capacidad de mostrar la realidad geográfica.

La IG es la referida a entes, elementos o fenómenos que ocurren sobre la corteza terrestre o en sus proximidades, en relación con los cuales es relevante el conocimiento de su posición referida a un sistema vinculado con la Tierra.

“Los entes, elementos geográficos o Unidades de Observación son fenómenos de interés para el mundo real que no pueden ser subdivididos en fenómenos del mismo tipo, pero pueden formar clases”.



Componentes de los elementos o entidades:

Todo elemento o fenómeno de interés se localiza en algún lugar (espacio), ha sucedido en algún momento (tiempo), y tiene asociada alguna medida del mismo (atributo):

Componente espacial

Geometría (localización): posición, forma, tamaño, orientación.

Topología (relaciones espaciales): deducible de la geometría.

Componente temática

Atributos o variables temáticas: son propiedades que pueden asignarse a cada localización del espacio, y que carecen de extensión específica; se estudian de modo genérico y están contenidos en códigos, nombre y atributos.

Componente temporal

Tiempo

Existen varias formas de hacer un análisis de los datos en función de sus componentes, las cuales se muestran a continuación:

Análisis espacial

Considera el aspecto espacial aisladamente

Estudia sus características geométricas puras

Análisis estadístico

Considera el aspecto temático aisladamente y utiliza las técnicas existentes.

Modelado cartográfico



Considera los dos aspectos simultáneamente e interactuando

Características de la IG

La IG tiene unas características muy particulares: es voluminosa, fractal, borrosa, dinámica y multiforme; a continuación describiremos cada una de estas características:

Voluminosa: Tanto por su procedencia y contenido muy dispar, el número potencial de datos en los trabajos de campo puede ser infinito.

Fractal: Los elementos naturales tienen formas sumamente irregulares, son objetos fractales. La dimensión fractal mide el grado de irregularidades. La representación depende de la escala del mapa; un mismo objeto varía sus dimensiones topológicas si se representa en mapas distintos; dependiendo de la escala, la presentación de la realidad es diferente.

Borrosa: Es difícil delimitar los elementos geográficos; por ejemplo, los límites de un bosque junto a un cultivo, el límite dónde acaba un río y comienza el mar, o entre una ciudad y un término municipal.

Dinámica: La información cambia con el tiempo y ello implica la corrección de los errores que se produzcan.

Cambios en el tiempo (mapas de uso en el suelo)

Cambios en el espacio (crecimiento urbano, obras hidráulicas, crecimiento demográfico (6% al año)).

Cambios temáticos (rotación de cultivos)

Multiforme: La representación de los objetos geográficos varía con la escala del mapa; una ciudad puede ser un punto o un polígono en función de la escala elegida. La elección dependerá del fin que se persiga y del tipo de las consultas a las que el sistema deberá responder.



Por otra parte, no cabe asimilar "información geográfica" a la mera localización de sucesos. La ubicación de fenómenos es frecuente en la indagación científica y en los modestos menesteres cotidianos; lo peculiar en aquella es el cometido estructurador de la posición, eje articulador de todos los análisis e interpretaciones. Precisamente, uno de los rasgos distintivos de la información geográfica es su riqueza y complejidad; en efecto, en una unidad elemental de información geográfica (*datum*) se cristalizan los siguientes tipos de información:

1. **Locacional**, para describir la posición de un objeto sobre la superficie terrestre mediante un sistema de coordenadas (x, y, z)
2. **Temática**, para expresar los atributos o propiedades de un objeto o localización
3. **Temporal**, para describir los cambios ocurridos con el transcurso del tiempo (este tipo de información puede omitirse en función de los objetivos de las aplicaciones)

El objetivo y desafío de los SIG consiste en representar la información geográfica con tecnología digital, de forma consistente y eficiente, y considerando al mismo tiempo otras características fundamentales de la misma como:

Tratamiento y análisis de la IG en un SIG

El tratamiento y análisis de la IG en un SIG implica abstracción y discretización:

Abstracción: estructuración de la información geográfica en capas temáticas.

Discretización: elección de determinados elementos geográficos, obviando otros que pueden no tener la importancia deseada.

Se debe tener en cuenta que los datos geográficos deben descomponerse en tres componentes: Temática, Espacial y Temporal, que responden respectivamente a las preguntas de ¿qué?, ¿dónde?, y ¿cuándo?



El SIG gira alrededor del interés por conocer cierto *aspecto del mundo real* y de intentar explicarlo mediante un *modelo*. Estudia el concepto de *medida* de los componentes como parte fundamental del desarrollo del modelo, y el concepto de *control* como generador del modelo del SIG.

Cada una de las medidas de los componentes juega un determinado papel en un SIG (Sinton estableció en 1979 que una debía ser “fija”, otra serviría de “control” y una tercera se mediría constituyendo la componente principal del estudio).

El control de los componentes da como resultado los marcos de medida que en los modelos SIG tradicionales son:

Atributos como control -> Modelo Vector

Espacio como control -> Modelo Raster

También existen en la actualidad un modelo que integra conjuntamente los dos modelos anteriores llamado Geo-objetos y que se sale del objetivo de este documento. En el siguiente apartado entramos a explicar en profundidad los modelos anteriores.

Modelos de datos de información geográfica

No existe acuerdo unánime sobre la clasificación de los modelos de datos de los SIG. Persiste cierta ambigüedad entre los conceptos modelo de datos y estructura de datos, sobre todo cuando un mismo adjetivo puede acompañar a ambos como sucede en modelo de datos topológico y estructura de datos topológica. Aunque las fronteras no son precisas, el concepto “modelo” es más abstracto que el de “estructura”. El mundo real ha de ser descrito en términos del modelo de datos, mientras que las estructuras de datos son esquemas operativos para gestionar información que se almacena físicamente en diversos formatos de registro. Por ejemplo, la altitud del terreno se puede representar con



un modelo de datos raster o vectorial; a su vez, cada modelo dispone de varias estructuras para alojar la información que se guarda en registros de variado formato (ver figura). Algunos autores sólo distinguen dos grandes grupos de modos de representación de la información geográfica, que dan nombre a los modelos: **vectorial** y **raster**; dentro del modelo vectorial, se dispone de varias estructuras como la TIN y la de red (*network*) para organizar los datos. Otros autores como M. Zeiler (1999) otorgan la categoría de modelo a la representación TIN, mientras que las redes se definen como una clase especial de relaciones entre elementos de un modelo vectorial; en cambio, para Longley, P., Goodchil, M., Maguire, D. y Rhind, D. (2001): *Geographic...*, p.186, la organización de datos en redes lógicas es un modelo diferenciado.

Modelo	Ventajas	Desventajas
Raster	Estructura de datos simple Compatibilidad imágenes de satélite y de escáner Buenas capacidades para el análisis, las simulaciones y el modelado Tecnología barata	Archivos muy grandes que se incrementan geoméricamente cuando aumenta la resolución Necesidad de estructuras de compresión de datos Menor precisión locacional
Vectorial	Buena representación cartográfica Estructura de datos compacta Facilidad de actualización de la base geográfica	Estructuras de datos complejas Imposibilidad para realizar ciertos análisis Tecnología cara



Adquisición de fuentes de datos

Estas fuentes de datos son fuentes realizadas por otras entidades o empresas y que nosotros utilizamos para sacar cierta información relevante. Existen dos tipos de fuentes principales:

Fuentes Documentales: este tipo de fuentes se basan en datos de censos, en registros, encuestas, etc. Por ejemplo, para el estudio de la población, unas fuentes típicas serían: el censo de la población, el registro civil, el padrón municipal de habitantes, etc.

Fuentes Cartográficas: se basan en la información de otras entidades, las cuales se encargan de generar elementos cartográficos. Estos elementos pueden ser Raster o Vectoriales. Destacamos entidades como el Instituto Geográfico Nacional, el cual tiene mapas de España (1:1000000, 1:200000), provinciales (1:200000), y autonómicos (1:100000, 1:400000). Otra entidad importante a nivel de nuestro país es el centro geográfico del ejército (CGN). El CGN ofrece tanto mapas sobre papel a distintas escalas como mapas digitales de tipo Raster o Vectoriales.

Los SIG se definen como tecnologías de integración de información de naturaleza diversa, por lo que es frecuente que cada proyecto utilice distintas fuentes de datos que a su vez se encuentran en formatos variados. Una clasificación simple permite distinguir entre fuentes primarias y secundarias y formatos digitales y no digitales.



Fuentes de información en los SIG

Documentos Fuentes	Digitales	No digitales
Primarias	<ul style="list-style-type: none"> - Levantamientos topográficos (vectorial) - Telemetría (vectorial) - Medidas de GPS (vectorial) - Imágenes de satélite (raster) - Otros instrumentos digitales 	<ul style="list-style-type: none"> - Observaciones de campo - Documentos lineales (Censos, encuestas, listas...) - Mapas - Fotografía aérea
Secundarias	<ul style="list-style-type: none"> - Imágenes de satélite (raster) - Bases de datos digitales - Listas (de direcciones, coordenadas...) - Documentos de escáner (raster) 	<ul style="list-style-type: none"> - Documentos de archivo - Otros mapas e imágenes

Los **datos primarios** suelen ser más congruentes con los objetivos de los proyectos, pero son costosos de obtener; a cambio, el investigador tiene más control sobre los mismos en lo referente a los individuos y unidades de medida y a los procesos de medición. El nuevo instrumental científico ha permitido mejorar la precisión y permanencia de los datos, al tiempo que facilita su recogida en formato digital.

Los **datos secundarios** han sido producidos por otros agentes, seguramente para otras finalidades, y medidos sobre individuos diferentes a los utilizados en un proyecto SIG. Este tipo de información se está incrementando considerablemente debido al crecimiento de



su demanda para realizar “mapas de consumo” (itinerarios, guías turísticas, etc.). Su utilidad depende de la escala y del modelo y estructuras de datos en que es publicada. A pequeña escala existe información para todo el planeta: bases GTOPO, bases creadas por instituciones cartográficas, publicaciones de empresas privadas, etc.

Captura de datos

La captura de datos es una de las fases más importantes de los sistemas SIG, ya que es la base de todo el sistema GIS. Esta fase se puede considerar bastante tediosa, costosa y larga, pero es lo primero que se necesita para conseguir un sistema GIS.

Se puede considerar que esta parte representa entre un 15% y un 60% de un sistema GIS. En cuanto a la técnica, no es muy difícil pero sí muy costosa, y además como parte del sistema GIS se puede considerar como una parte bastante crítica, ya que si la captura de los datos no ha sido la correcta esto generará muchos errores y luego es muy costoso el trabajo para mejorar los datos conseguidos.

Existen dos métodos de captura de datos: los métodos primarios y los métodos secundarios. Los métodos primarios constan de una serie de métodos vectoriales (levantamientos, GPS, fonometría) y otros métodos raster (teledetección, ortofotografías). Estos métodos son aplicados específicamente para los sistemas GIS, dando una mayor adecuación de los datos y una calidad más uniforme. Para la generación de estos métodos primarios se necesitan unos requisitos básicos:

Los procesos necesarios han de ser automatizables.

Los datos generados deben salir en formato digital.

Han de ser rápidos y de fácil uso.



Deben posibilitar la captura de información asociada a los elementos.

Deben permitir la creación de topología.

Los problemas que generan estos métodos primarios son:

Se introducen errores debido a la automatización.

Dificultad de aprendizaje.

Problemas más propios de las técnicas empleadas que del SIG propiamente dicho.

Los métodos secundarios constan de una serie de métodos vectoriales (digitalización, BD alfanuméricas y Datos CAO) y un método raster (escaneado). Estos métodos secundarios no han sido creados explícitamente para los sistemas GIS, pero a partir de ellos se puede utilizar la información para los SIG. Esto hace que los datos conseguidos no sean mejores que los primarios.

Una vez conocidos los métodos existentes, se tienen que diseñar unas fases para la captura de datos. Estas fases han de tener una estructura en común. La estructura es la siguiente:

Diseño de especificaciones: en esta fase se debe decidir qué datos se necesitan para nuestro sistema. Aquí se debe analizar el porqué y el para qué de estos datos.

Preparación: la zona que se quiere elegir se debe preparar sobre mapas de papel para así poder minimizar los errores.

Captura y estructuración: lo que se va a obtener en una primera instancia son una multitud de puntos, y a partir de ellos se debe elegir la forma de dibujar los elementos gráficos (directamente, o mediante CAD)

Validación: se deben generar métodos sencillos para validar los datos de la captura.



Utilidades, campos de aplicación y aplicaciones.

En general, las utilidades más extendidas de los GIS son las siguientes:

Presentaciones y cartografía temática: el SIG representa datos sobre un mapa para mostrar localizaciones de diferentes actividades o características, o para representar sus atributos. Son los mapas temáticos.

Consulta de datos: todas las empresas recogen y mantienen extensas bases de datos, la mayoría con referencias espaciales. Requieren de tablas y listados. El SIG muestra todo esto por medio de un mapa. Es la técnica “*show me*” (muéstrame).

Consulta espacial: el mapa, como herramienta de consulta, permite el acceso a la base de datos señalando una zona concreta y presentando los resultados en una tabla o resumen. Es la técnica “*tell me*” (dime).

Integración y actualización de bases de datos: una de las formas de llevar la organización de las bases de datos es por la localización. El SIG es un sistema versátil y flexible para introducir, mantener, actualizar y consultar bases de datos. La clave es el “*geocoding*”, que es la capacidad para asociar el acontecimiento, la actividad o la observación a una característica espacial.

Cálculo y optimización de rutas: muchas veces se quiere encontrar el camino más corto a lo largo de una red de transportes. Este recorrido debe ser “legal”, es decir que los giros o transbordos sean correctos. Se busca la ruta más corta, más rápida y más apropiada entre un conjunto de alternativas. Los algoritmos de cálculo y optimización de rutas del mundo real permiten que el SIG disponga de esta propiedad.



Análisis de distancias, adyacencia y proximidad: el SIG permite calcular distancias sobre un mapa, en línea recta o a lo largo de una red. Además, puede determinar las características de mapas adyacentes o cercanos.

Campos de aplicación

Las aplicaciones que presentan los SIG son numerosas, pero se podrían clasificar en tres grandes bloques, los dos primeros basados en campos y el tercero en objetos:

Aplicaciones cartográficas: irrigación, análisis del rendimiento de las cosechas, evaluación del terreno, planificación y gestión de instalaciones, estudios del paisaje, o análisis de la estructura del tráfico.

Aplicaciones para el modelado digital del terreno: estudio de recursos de las ciencias de la tierra, análisis del suelo, estudios de la contaminación del aire y del agua, o control de inundaciones, entre otras.

Aplicaciones de objetos geográficos: sistemas de navegación de vehículos, análisis del mercado geográfico, distribución y consumo de servicios públicos, o análisis económico de productos y servicios para el consumidor.

Los sectores en lo que se pueden implantar los sistemas SIG son muchísimos. A continuación mostraremos un conjunto de sectores:

Sector bancario: localización de red de sucursales en función de las características de la población. Estudio de modelos de mercado potenciales. Estudio de riesgos en la gestión de seguros. Seguimiento de inversiones y de los resultados del mercado bancario en su dimensión territorial.

Sector de estudios de mercado: segmentaciones de mercado, distribución territorial de la población y de sus características socioeconómicas.



Sector sanitario: seguimiento de estudios epidemiológicos. Planificación de la red de asistencia sanitaria en relación con la población que habita una zona. Análisis de la distribución geográfica de los perfiles sanitarios de la población.

Sector logístico: gestión de flota. Planificación y optimización de rutas. Determinación de centros de distribución. Posicionamiento de puntos de venta y análisis de itinerarios de recogida y suministros. Control de los envíos.

Sector de las telecomunicaciones: planificación de las redes de telefonía móvil, de televisión por cable, análisis de cobertura del medio.

Sector de la comunicación: información sectorial para su transmisión gráfica hacia los sectores deseados de la opinión pública. Análisis de los efectos de las campañas de publicidad y promoción. Teletrabajo, educación a distancia, tiempo libre, información sobre ocio.

Sector de franquicias: localización de nuevos puntos de venta, captación de clientes potenciales.

Sector del automóvil: de los 14,5 millones de automóviles de turismo que se matricularon en Europa en el año 2000, cerca del 47% ofrecían la opción de contar con un navegador. En muy pocos años será una necesidad.

Sector medioambiental: para realizar inventarios de suelos o controlar el tipo del uso del mismo.



Área de aplicación	Aplicaciones
Gestión de recursos naturales y del medio ambiente	Seguimiento de la contaminación, del tiempo y clima; evaluación de impactos ambientales; inventario de recursos; biodiversidad y conservación; estudios sobre el paisaje; planificación hídrica (ríos, acuíferos, zonas de inundación); espacios protegidos; espacios naturales (control de incendios, de plagas)
Agricultura	Seguimiento de cosechas, control de plagas y enfermedades; estudios de suelos; planificación de riegos y usos del agua
Comercio y servicios	Análisis de localización de establecimientos; estudios de oferta/demanda; publicidad; evaluación del potencial de mercado, áreas de influencia; accesibilidad
Transporte	Planeamiento de infraestructuras (carreteras, ferrocarril, eléctricas, telefónicas, abastecimiento de agua), control de vehículos), determinación de rutas óptimas; control de los flujos de las redes (agua, electricidad)
Servicios públicos	Estudios epidemiológicos; análisis de accidentalidad; control y seguimiento de rutas de ambulancias y vehículos públicos, planificación espacial de los centros de salud, docentes, administrativos, etc.
Gestión local	Protección civil (bomberos, policía, actividades molestas y peligrosas, planes de evacuación, etc.); seguimiento y control de licencias de obras; catastro; mantenimiento de infraestructuras (calles, semáforos, etc.); planificación y control (usos del suelo, zonificaciones, etc.); planificación y seguimiento de la política ambiental



Herramientas de apoyo para la toma de decisiones

El primer factor es el grado con que las funcionalidades ofrecidas por el SIG corresponden al tipo de operaciones que se le exigen. Existe actualmente un mercado sustancial para las aplicaciones especializadas; en algunos casos, las herramientas específicas tienen que ser agregadas a las existentes. Esta demanda indica que aún existe un vacío entre las necesidades del usuario y lo que los software de SIG pueden ofrecer. Por otro lado, no se puede poner el software de SIG en la misma categoría de, por ejemplo, el software de procesamiento de palabras. Los diferentes tipos de aplicaciones de los Sistemas de Información Geográfica requieren utilidades altamente específicas que no pueden ser cubiertas a través de un solo paquete de software.

El segundo factor es determinar hasta qué punto el software es fácil de usar, por ejemplo, a través de una interfase entre el usuario y el software. ¿Pueden los usuarios de los SIG usar el software sin ayuda, o se necesita algún tipo de apoyo permanente? En años recientes, se han agregado todo tipo de herramientas a los software de SIG, permitiendo a los usuarios construir su propia interfase especializada.

El tercer factor se refiere a las inversiones necesarias en software y hardware. La disponibilidad de muchas herramientas de software de SIG en PCs y el bajo precio del hardware (PC y estaciones de trabajo) indican que el costo de hardware y software no es un gran obstáculo. La educación y el conocimiento constituyen el cuarto factor. Como cada vez las personas y disciplinas se involucran más con los SIG, la falta de conocimiento de los mismos es cada vez menos un problema; sin embargo, no todos son conscientes todavía del enorme potencial de los SIG para los negocios.



Igualmente, el número de personas que han sido o están siendo entrenadas en SIG está creciendo. El problema es más una cuestión de calidad que de cantidad. No se ha podido determinar si los conocimientos de los SIG adquiridos en las universidades, institutos y centros de capacitación y en general en todo tipo de cursos, satisfacen la demanda.

El quinto factor es el problema de los datos. Este es un problema mayor porque disminuye la velocidad del proceso de difusión del uso de los SIG. Las inversiones en datos son altas y los problemas relacionados con la disponibilidad, el costo, los estándares, la exactitud y las obligaciones legales están lejos de resolverse. Debido a las actividades de recolección de datos por parte del sector gubernamental y por iniciativa privada, la disponibilidad de datos ya no es un problema tan agudo. En cambio, el costo de los datos es ahora el factor que más dificulta el uso de información geográfica, aunque fundamentalmente los problemas de estándares, exactitud, y obligación legal en el uso de SIG son tomados en cuenta una vez que el problema del costo se ha superado. Se espera que el problema de los datos siga siendo el factor más importante en el éxito comercial en los próximos años.

Los SIG en los negocios

En general, el uso de los SIG surgió de la necesidad de crear y mantener grandes bases de datos espaciales y de la necesidad de hacer cartografía. Grimshaw (1993) describe las aplicaciones comerciales de los SIG como una herramienta para apoyar la toma de decisiones.

La mayoría de los gerentes manejan datos con una dimensión geográfica; el uso de los SIG les permite identificar un patrón espacial en sus datos, hasta ahora desconocido. La



importancia de los SIG para las aplicaciones comerciales es obvia: conocer dónde están los mercados potenciales es crucial para cualquier negocio.

Las aplicaciones operacionales incluyen, por ejemplo, el uso de las funcionalidades de los SIG para supervisar la provisión de productos en una red de distribución. Para este propósito, los SIG serán utilizados para apoyar actividades diarias de rutina. Las aplicaciones tácticas proporcionan información requerida para la toma de decisiones. Estas aplicaciones tácticas generalmente son usadas por la gerencia media. El proceso de toma de decisiones requiere la combinación de todos los tipos de datos pertinentes a la decisión. Los SIG apoyarán la combinación de datos espaciales: por ejemplo, información sobre mercado potencial y dónde se localizan los competidores, para decidir la ubicación de un nuevo canal de distribución; además, el uso de los SIG tácticos también puede orientarse a problemas como dónde dirigir una campaña publicitaria para un target específico.

Las aplicaciones estratégicas de los SIG están dirigidas a la alta gerencia. El sistema tiene que proporcionar la información ad hoc que se necesita para tomar decisiones estratégicas. La alta gerencia se apoyará en la toma de decisiones estratégicas gracias a la facilidad de cartografiar de un SIG.

Problemas que podrían involucrar a la alta gerencia son, por ejemplo, dónde invertir o qué nuevos productos lanzar al mercado. Esta es simplemente una manera de clasificar las aplicaciones de los SIG en los negocios. No obstante, enfatiza un punto importante: estar consciente que cada nivel dentro de una organización y cada tipo de actividad requiere un tipo diferente de SIG. Esto también está relacionado con el proceso de difusión de la



tecnología de los SIG, porque hay una clara secuencia de difusión según el tipo de aplicación.

En general, las aplicaciones operacionales de los SIG serían llevadas a cabo primero en una organización. Para tres tipos de operaciones fue fácil mostrar que los SIG pueden ayudar a ahorrar dinero. Además de estos efectos directos, el uso operacional de los SIG también estimuló el conocimiento de los SIG dentro de una organización. Esto habilita el próximo paso de este modo: el uso táctico de los SIG. Hoy en día se dedican muchos esfuerzos a aplicar los SIG a los problemas tácticos. Es esencial integrar bases de datos internas y externas para un óptimo proceso de toma de decisiones. El grado en que la alta gerencia estará usando un SIG para consultar información espacial en sus decisiones estratégicas dependerá del éxito en la implementación de las aplicaciones operacionales y tácticas en su organización. Al mismo tiempo, dependerá de qué tan fácil sea de usar es el software.

5. Diseño, Organización e Implementación de un SIG

Introducción.

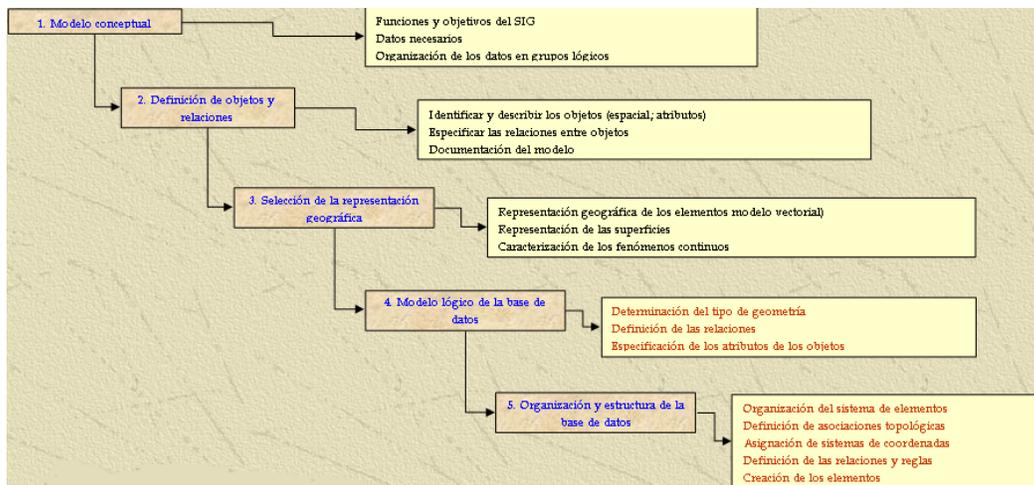
Las empresas del sector privado o público que utilizan información espacial pueden considerar, o pueden haber sido aconsejadas, que la tecnología SIG mejoraría la eficiencia de sus operaciones. Sin embargo, se plantea el problema de si esto es cierto o no, y si dicho sistema debería ser desarrollado específicamente para la empresa en cuestión, o adquirido mediante un sistema ya existente. La complejidad de la tecnología SIG en la actualidad, y sus aplicaciones, hacen que no sea corriente el desarrollo propio como lo fue en los primeros momentos de desarrollo de los sistemas de información geográficos. El tiempo, los costos y los riesgos empleados en el desarrollo y mantenimiento de los



grandes sistemas informáticos, combinados con la disponibilidad actual de software comercial de GIS, hacen que la mayoría de las empresas se incline hacia la adquisición de un sistema ya existente, adaptándolo a sus necesidades en vez de desarrollar uno propio.

La mayoría de los sistemas comerciales presentan una amplia gama de posibilidades funcionales, de configuraciones del sistema, estructuras de datos, características de funcionamiento y costos. Sin embargo, hay poca literatura sobre el asunto en comparación con otros sistemas informáticos, y sus características cambian tan rápidamente que pronto se queda desfasada en el tiempo. Se requiere por consiguiente una metodología para cualquier empresa, para determinar si el beneficio de la adquisición de un SIG es mayor que sus costos, y si esto es así, para decidir qué sistema adquirir.

Podemos encontrarnos dentro del universo de un SIG con una serie de elementos que se han de considerar. Un SIG no son sólo los ordenadores, programas y aplicaciones informáticas, sino que está compuesto también por un personal que, además de diseñarlo, se encarga de su implantación, y de un personal que trabaja con él. No se puede olvidar la metodología de implantación, que en buena medida va a condicionar los resultados del mismo. Por último, pero no por importancia, nos encontraríamos con los propios datos que gestiona el sistema, sin los cuales no podría funcionar.



Aunque la definición de fases y su nomenclatura varían de un autor a otro, se pueden distinguir una serie de elementos comunes a todos los modelos de adquisición de SIG. Siguiendo un esquema general (según A.L. Clarke), común a la metodología de otros sistemas, tendríamos como fases:

Fase 1: Análisis de requisitos.

1. Definición de objetivos.
2. Análisis de necesidades del usuario.
3. Diseño preliminar.
4. Análisis de costo-beneficio.
5. Estudio piloto.

Fase 2: Especificación de requisitos.

1. Diseño Final.
2. Solicitud de ofertas.



Fase 3: Evaluación de alternativas.

1. Listado breve.
2. Prueba de control (benchmark test).
3. Evaluación costo-efectividad.

Fase 4: Implementación del sistema.

1. Plan de implementación.
2. Contrato.
3. Test de aceptación.
4. Implementación.

Para hacer el seguimiento de las fases recurrimos, sin embargo, al siguiente modelo simplificado y más funcional, al que se corresponden las siguientes partes:

1. Revisión de objetivos.
2. Análisis de requisitos.
3. Diseño conceptual.
4. Estrategia de implantación.
5. Implementación.
6. Puesta en marcha.

Aunque coinciden en diversos puntos, difieren a la hora de ubicar los estudios y divisiones de las fases, por lo que en lo sucesivo nos referiremos a este modelo. En dicho modelo, además de la metodología por fases, se incluyen las actividades y tareas a realizar dentro de cada una de las fases, así como los diversos interlocutores con los que se ha de contactar en cada momento dentro de la empresa, los diversos formularios que se han de



rellenar para los estudios en concreto, y los documentos resultado de cada una de las actuaciones que se van realizando.

Fase 1: Revisión de objetivos.

En esta fase se pretende delimitar y conocer el ámbito y los objetivos para el SIG. Es necesario realizar un estudio de oportunidades del SIG. En general, nos encontraríamos con una revisión de todos los objetivos de la empresa, el desarrollo de los objetivos para el proyecto SIG, y la discusión y negociación con la gestión de la empresa teniendo en cuenta a los usuarios. La adquisición del SIG ha de ser compatible con el propósito de la organización de la empresa al que va dirigido, y estar acorde con su plan de negocio. Ha de encajar en las líneas de actuación y con los objetivos estratégicos, no ir en contra de las iniciativas ya en marcha. Debe prever la interacción con sus clientes y otras organizaciones con las que se relaciona. Los aspectos principales deben incluir costo, tiempo, calidad, precisión y el impacto que tendrá sobre la plantilla de la empresa. Todo ello hará que sea necesario una revisión del plan de negocio.

Los interlocutores en este caso serán los gerentes y responsables de la empresa, dentro del nivel de decisión y de gestión. Necesitamos una serie de formularios en los que han de ir relatados los datos generales de la empresa, los datos de las distintas unidades de negocio que operan, el organigrama de la organización, los objetivos generales, los objetivos que se pretenden con el SIG, y los beneficios que aporta el SIG.

Como documento final de esta fase tendremos la propuesta de proyecto SIG, en la que se incluirán las funciones que desarrollará el SIG, sus costos, tiempos (de implementación, aprendizaje, etc.) y los beneficios que revertirán a la empresa.



Fase 2: Análisis de requisitos.

En esta fase se pretende determinar los requisitos necesarios para el usuario, sobre los que el SIG va a ser diseñado y evaluado. El primer paso es la preparación del propio análisis, en el cual se estudian los procesos y flujos de información actuales en la empresa y se identifican las necesidades de información y usuarios; esto es, qué información se usa y quién la usa.

A continuación se requiere una definición de productos de información, que es la clave del proceso. Los productos pueden estar en múltiples formatos, tanto físicos como de los propios formatos de salida de la empresa, y lo mismo ocurre con los datos de tipo digital. En primer lugar se confecciona una lista de necesidades. Igualmente, es necesaria una descripción de productos. Los productos tipo los encontraremos en la forma de mapas, informes y documentos.

Las aplicaciones del SIG y la captura de datos pueden requerir una variedad de productos intermedios que nos proporcionen verificación del proceso. Los productos que proporcionan la información geográfica en ese momento constituyen el punto de partida para determinar las necesidades de los usuarios, con la posibilidad de proporcionar productos nuevos y mejorados (para los que se ha creado la necesidad de incluir un SIG). Es necesario un estudio de los datos y funciones necesarias dentro del esquema del SIG. Los requisitos de los datos están determinados directamente por la definición de los productos. El análisis debe identificar las clasificaciones, la precisión y la frecuencia de actualización requeridas para cada tipo de datos. Las funciones han de cubrir las relaciones lógicas entre datos, concretamente los niveles entre mapa-atributo, atributo-atributo y mapa-mapa.



Dentro de la precisión nos encontramos con la tolerancia al error. Nos podemos encontrar con errores referenciales, de inequación entre elementos y referencias, como pueden ser los códigos de parcelas mal referenciados. Existen así mismo errores topológicos, como pueden ser redes desconectadas o nudos colgados. Por último, tendríamos errores relativos o absolutos, como pueden ser las distancias entre elementos o la discordancia entre conjuntos de datos.

Todo el análisis repercute en un resumen económico. Hay que valorar los costos actuales en el uso y generación de datos y ver los posibles beneficios generados al reemplazar los sistemas actuales por el SIG propuesto.

Por último, se hace una determinación del alcance que tendrá la implantación. Se fijan en esta parte las prioridades y los recursos disponibles en la actualidad. En las prioridades tendremos una enumeración de las funciones tipo a realizar y el número de veces que se utilizará cada una. De esta manera se podrán clasificar como de alta, media o baja prioridad. Las de alta prioridad deberán ser imprescindibles dentro del sistema. Las de media son deseables, pero su desarrollo no ha de ser tan exhaustivo como las primeras. Las últimas tendrán una aparición esporádica y como tales se tratarán. Las características de funcionamiento y el marco de trabajo del SIG tendrán un gran impacto en la configuración del hardware propuesto, y por consiguiente en los costos. Aspectos importantes son el número simultáneo de usuarios, los volúmenes de datos usados, los tiempos de respuesta y las necesidades de producción.

El análisis de requisitos del usuario puede llevar a una mejora en la definición de objetivos. Los interlocutores que se van a encontrar en esta fase van a ser, dentro del nivel de gestión, los jefes de área o servicio; y a nivel operativo, los usuarios expertos que en definitiva utilizarán el sistema y conocen la estructura y los usos de los datos.



Para conseguir toda esta información se recurre a una serie de formularios. Vendrán reflejados en ellos los procesos actuales y la arquitectura de la información. Se deberán identificar los usuarios potenciales del SIG. Nos encontraremos con la lista de necesidades, la descripción de productos de información, la lista maestra de datos de entrada, las relaciones lógicas, las tolerancias al error, así como la lista maestra de funciones y su frecuencia de uso. En definitiva, habrán de estar reflejados todos los aspectos del análisis citado anteriormente.

Los documentos resultantes de esta fase serán el documento de requisitos del SIG y un resumen ejecutivo.

Fase 3: Diseño Conceptual.

El objetivo en este caso es estructurar y delimitar el contenido del sistema. La información recogida en el paso anterior permite hacer un diseño preliminar para el desarrollo del SIG. Las actividades a desarrollar en este caso serán un diseño conceptual de la Base de Datos a utilizar, y otro de las funciones.

Dentro del diseño de la base de datos habrá que tener en cuenta los requisitos de los datos, así como su precisión y los estándares que se van a utilizar. Se deberán integrar los datos existentes. En el manejo de estos datos dentro de la base de datos tendremos unas opciones de disponibilidad y de necesidades de conversión entre tipos de datos. La clasificación, la precisión y la frecuencia de actualización para cada tipo de dato ya fueron identificadas durante el análisis de los requisitos de los datos. Las especificaciones de la base de datos deben también identificar las fuentes, volúmenes y estructuras tanto para los datos espaciales como para los atributos. Estas consideraciones previas deberán conducirnos a la elección de un modelo espacial de tipo raster o vector (o ambos), y a



escoger entre tipo relacional pleno u otro tipo de base de datos para los atributos (en definitiva, a escoger el modelo conceptual de datos).

Las especificaciones del diseño de las funciones del SIG están determinadas directamente por la definición de los anteriores productos y las especificaciones de la base de datos. Estos definen las funciones y procesos que se requieren para permitir el desarrollo de la base de datos y la producción de productos de información; en definitiva, un análisis de utilización de las funciones y sus propiedades. Guptill (1988) propone una lista de funciones genéricas de SIG para interfaces de usuario, manejo de la base de datos, creación de la misma, análisis y manipulación de datos, presentación de los datos y generación de productos. Según los requisitos de interfaz, comunicaciones (incluidas las inter-funcionales), hardware y software, nos encontraremos con unas u otras funciones. En definitiva, lo que se presentará al final del proceso es una serie de ventanas al usuario con unas funciones asociadas e interrelacionadas entre sí, con salidas en pantalla o a través de impresora u otro dispositivo de salida.

Los modelos conceptuales deben ser desarrollados y documentados para describir el diseño lógico y físico del sistema propuesto. Deben incluir los aspectos relativos al hardware, software, comunicaciones, procesos, recursos humanos y las líneas de acción de la organización. Se pueden incluir modelos alternativos para la arquitectura del hardware y de las comunicaciones. Para algunas empresas, la elección entre un modelo centralizado o distribuido del SIG puede suponer un elemento crítico del diseño. Dangermond (1988) recoge las tendencias, ventajas y desventajas de varias arquitecturas de SIG.



Como en las otras fases, se deberá contactar con una serie de interlocutores válidos para el desarrollo del diseño. En este caso se contará con los técnicos de la organización, dentro del nivel operativo.

Para llegar a este diseño preliminar contamos con formularios donde se requerirá la especificación de la base de datos, la especificación funcional del sistema y el diccionario de datos. Tendremos como resultado final un documento que contendrá el diseño preliminar del SIG en los dos aspectos mencionados.

Fase 4: Estrategia de implantación.

En esta fase se plantea cómo llevar a cabo la implementación del SIG. Varios son los aspectos a tener en cuenta en este momento. Nos encontraríamos con la estrategia a seguir y el análisis costo-beneficio.

Para definir la estrategia debemos recurrir al estudio de diversas variables. La tecnología existente supone un primer factor de estudio. Un estudio de mercado informal tras el diseño preliminar debería estar dirigido a determinar las capacidades del sistema. En un principio se podría realizar mediante visitas a vendedores y a puestos de usuarios. El objetivo sería determinar si las especificaciones preliminares pueden ser suplidas por la tecnología existente. Si esto no es así, las opciones serían rebajar la funcionalidad y las especificaciones de uso del sistema, o aceptar la posible adquisición de un sistema de mayores componentes de desarrollo.

Otro de los elementos sería un estudio de las aplicaciones del SIG en relación con los datos. A medida que se realice la implementación del SIG deberán articularse las diversas fases para que el sistema, lo antes posible, empiece a generar los primeros resultados a través de sus aplicaciones. Para ello se recurre a la elaboración de cronogramas que indican la distribución de las diversas actividades en el tiempo.



Dentro de la estrategia no se puede olvidar el factor humano. El personal de la empresa necesitará adaptarse paulatinamente al nuevo sistema. Para ello será necesario un periodo de formación del mismo que se acomode al proceso de implementación, de manera que el sistema esté operativo sin demoras debido a la falta de manejo de los usuarios.

Por último, habrá de realizarse un estudio de riesgos-seguridad. Se evaluará en tanto por ciento según los posibles factores de riesgo, aunque éste presente dificultades, entre otras por los múltiples interlocutores, sobre todo por la toma de decisiones y las distintas prioridades.

El análisis costo-beneficio tiene por objeto establecer los argumentos económicos para la propuesta de adquisición del SIG. Los costos, beneficios, impactos y riesgos de la adquisición del sistema se miden en relación con los que se tendrían al mantener los productos, datos y procesos actuales de la empresa. Si el diseño preliminar admite diversas alternativas, será necesario analizar los costos y beneficios de cada una de ellas. Las fases del análisis consisten en estimar el modelo de costos completo del SIG, valorar los beneficios, valorar los riesgos y el cálculo del beneficio neto ponderado.

Los costos de implementación y operación del SIG incluyen los de adquisición y mantenimiento del hardware y software, captura y almacenamiento de datos, formación de personal adicional y más calificado (requerido para el manejo, programación propia, soporte a los usuarios y ejecución de aplicaciones del sistema), consumibles, preparación de locales y todos los gastos generales asociados. El costo del proceso de adquisición puede ser incluido también. Los costos constantes deberían determinarse a partir de un modelo de vida del sistema en el tiempo, de al menos cinco años. El descuento refleja el



costo de oportunidad del capital y permite la comparación de costos y beneficios que ocurran en diferentes tiempos durante la vida del sistema.

Se podrían obtener unos costos indicativos del hardware, software, mantenimiento y formación a través de dos o tres de los vendedores identificados anteriormente. Si fuera posible, estos costos se podrían validar a través de contactos con otras empresas usuarias ya existentes. Los costos de captura de datos pueden oscilar entre 10 a 1000 veces el costo del hardware y el software. Los modelos del proceso de digitalización manual desarrollados por Marbie, Lauzon y McGranaghan (1984) y Lai (1988) proporcionan una base para evaluar este componente.

Se pueden definir tres categorías de beneficios de un SIG: de eficiencia, de eficacia e intangibles (Prisley y Mead 1987; Maffini y Saxton 1987). Los beneficios de eficiencia son los relativos a los ahorros de tiempo y costo debidos a un procesamiento de datos más rápido y a la reducción de resultados duplicados, mientras que los beneficios de eficacia se refieren a la mejora de los procesos de toma de decisiones debidos a la nueva distribución temporal o a la información nueva. Los beneficios intangibles pueden incluir la mejora de la imagen pública de la empresa, la reducción de la confusión ante datos contradictorios, la mejora de la cooperación entre los usuarios a través de compartir los datos, el incremento de la profesionalidad y la moral del personal, la mejora de la posibilidad de hacer frente a imprevistos, y los nuevos conocimientos debidos a la mejora del análisis de los datos.

Mientras que es fácil asignar un valor económico a los beneficios de eficiencia, los otros dos beneficios son más difícilmente cuantificables. Dickinson y Calkins (1988) y Dickinson (1989) describen métodos para estimar el valor económico de los beneficios no



cuantificables. Sin embargo, Dickinson y Calkins (1988) también recomiendan que tales valores sean presentados separadamente ya que tienen una gran componente de incertidumbre. Chorley (1988) argumenta que en lugar de cuantificar beneficios que son inherentemente intangibles, es mejor proporcionar un análisis y descripción claros de tales beneficios para permitir la evaluación por analistas experimentados.

Los impactos del SIG propuesto sobre la organización y la plantilla de la empresa pueden ser grandes y también pueden tener una relación significativa con el análisis costo-beneficio. Los impactos en la recopilación y procesamiento de datos y en el personal de toma de decisiones pueden ser valorados en el análisis de requerimientos del usuario. Los impactos en la organización pueden incluir cambios en la estructura de organización y gestión asociados con las nuevas tecnologías, nuevos papeles y procedimientos, y nuevos requerimientos de consulta y cooperación. Estos elementos institucionales pueden tener mayor influencia en el éxito del SIG que los técnicos (Foley 1988; Seldon 1987). Una reunión temprana con el personal y sus representantes, advirtiendo sobre estos impactos, ayudará a evitar conflictos durante la implementación. Puede haber implicaciones legales y políticas para la empresa en términos de responsabilidad, autoridad y garantías con respecto a la recopilación y mantenimiento de los datos y la diseminación de productos de información.

El análisis costo-beneficio debe incluir una valoración del riesgo de que el proyecto no consiga un resultado próspero en términos de tiempo, costo, especificaciones y beneficios. El riesgo económico puede ser valorado mediante un análisis delicado, determinando los valores más optimistas y pesimistas para los costos y beneficios cuantificados. Otros factores de riesgo incluyen la complejidad de los datos y del sistema



considerado, la experiencia y composición del equipo de proyecto del SIG, y el impacto previsto del sistema en la organización y plantilla. La descripción de los riesgos en este momento inicial permite actuar a los gestores para reducir los riesgos, o seguirlos de cerca durante el proyecto.

Los costos, beneficios, impactos y riesgos pueden ser analizados y presentados de varias maneras. Dickinson y Calkins (1988) describen un modelo de costo-beneficio basado en los valores del producto SIG. El requisito mínimo es presentar los aspectos básicos económicos de la propuesta junto con una relación de factores no incluidos en el análisis económico. La ecuación económica básica es:

$$\text{Costo operativo de reemplazo del sistema} - \text{costo operativo y de capital del SIG propuesto} + \text{beneficios cuantificados del SIG propuesto} = \text{beneficio económico neto del SIG propuesto}$$

Este beneficio económico neto debe ser valorado a continuación con respecto a informes concisos de los beneficios intangibles y de los riesgos, así como de los impactos en la organización y en el personal. La consideración del análisis costo-beneficio es un hito principal en el proyecto de adquisición del SIG. Puede mostrar que la propuesta debe ser aplazada, que se necesita un trabajo mayor a la hora de fijar los objetivos, los requerimientos y el diseño preliminar, o que la adquisición pase a la fase siguiente.

Clarke incluye en esta fase, a continuación del análisis costo-beneficio, un estudio piloto para examinar el diseño preliminar antes de llegar al diseño final e invertir más recursos, así como para desarrollar la comprensión y la confianza de los usuarios en la nueva tecnología, y ganar algo de experiencia operacional para ayudar al diseño de lo que será la



prueba de control (*benchmark test*). El estudio piloto comprende como actividades su propio diseño, la selección del sistema piloto, la adquisición de los datos para el mismo, la producción de resultados y su análisis. El documento de diseño del estudio debe hacer constar los objetivos del estudio, dirigir la selección del sistema piloto, de los datos y productos, e identificar los recursos requeridos y el horario propuesto. Es importante para los usuarios estar de acuerdo con los objetivos y el alcance del estudio.

Durante el diseño preliminar se habrán identificado un número de sistemas potenciales. La selección de uno de ellos para este estudio se basa en el ajuste aparente de sus capacidades a los requerimientos, y al costo de establecimiento del estudio piloto. El hardware y el software pueden ser arrendados durante el estudio, o los propios proveedores del sistema pueden prestar apoyo. Ha de indicarse claramente a todas las partes que la elección de un sistema para el estudio no implica que sea el finalmente adquirido. Los usuarios deben evitar el peligro de sentirse obligados a escoger el sistema por su familiaridad con él.

Los datos piloto deben incluir ejemplos de todos los tipos de datos especificados en el diseño preliminar. Un acercamiento al problema es escoger un área geográfica pequeña pero representativa y adquirir todos los datos del área. Los productos piloto deben ser por consiguiente representativos del sistema final, e incluir aquellos que los usuarios consideren fundamentales para el éxito del sistema. Si hay conflicto entre escoger datos tipo raster o vectorial, los datos y productos piloto deben ser seleccionados para resolver esta cuestión.

El estudio piloto proporcionaría una experiencia valorable, así como sugerencias de los usuarios. Los resultados deben conducir a la mejora de la base de datos y de la



especificación del sistema, y a revisar los elementos del costo y beneficio y la relación de beneficios intangibles, impactos y riesgos.

En toda esta fase los interlocutores dentro de la empresa se encuentran en el nivel de gestión, siendo los jefes de área o de servicio.

Se cuenta como formularios todo aquello que interviene en el proceso, como la relación de los costos del SIG, la valoración de los beneficios, de los riesgos y del beneficio neto ponderado.

Como documentos finales tenemos el plan de implantación y un resumen ejecutivo.

Fase 5: Implementación.

Tras la aprobación definitiva de la estrategia de implantación, se llega en este punto de la metodología a la construcción final del SIG. Se puede estructurar esta fase con tres sub-fases, que incluirían el diseño final del sistema, la construcción propiamente dicha del mismo y la implantación como tal dentro de la empresa.

Diseño final.

El objetivo de este paso es producir la documentación del diseño necesaria para la adquisición. Para ello se tendría el diseño de la arquitectura del sistema. Los estudios anteriores del mercado habrán fijado los límites de la tecnología existentes para realizar el diseño detallado. Se finalizarán las especificaciones de la base de datos (tipos de datos, adquisición de los mismos, etc.), así como las especificaciones funcionales (las aplicaciones, el modelo de procesos, los módulos, etc.) y las genéricas del sistema; si se ha realizado un estudio piloto sus resultados se incorporaran al proceso.



En esta parte conviene insistir en la especificación detallada de las funciones y en su clasificación como obligatorias (corresponderían con la clasificación de prioridades), deseables u opcionales. Únicamente aquellas que sean esenciales deberán ser clasificadas como obligatorias, de lo contrario una sobre-especificación redundará en una difícil evaluación de alternativas y podría provocar la desaparición de propuestas de innovación. Las especificaciones de ejecución deben ser expuestas en términos de actuaciones mínimas aceptables (capacidad de trabajo obligatoria) y requerimientos óptimos.

Hay una serie de elementos previamente existentes que deben ser identificados y especificados, como son el hardware, el software, los sistemas de comunicación, los requisitos del interfaz y la política de la empresa sobre compatibilidad y estándares, ya que suponen condicionantes del sistema a los que, en lo posible, hay que adaptarse.

Los requisitos genéricos del sistema incluyen el mantenimiento, el soporte técnico, la formación, la documentación del usuario y del sistema, las herramientas del desarrollo, las actualizaciones, la seguridad y la ergonomía.

Los interlocutores válidos de todo este proceso se encuentran en el nivel operativo, ya que son los técnicos de sistemas los que mejor conocen las necesidades y el funcionamiento básico de la empresa.

El contacto con estos interlocutores deberá tener como resultado una serie de formularios que incluirán toda la serie de requisitos finales del SIG.

Al concluir este paso se tendrá el documento que refleje el diseño detallado del sistema.



Construcción.

El paso siguiente sería llevar a la realidad el diseño detallado obtenido. La primera operación a realizar sería la adquisición de la tecnología necesaria. Para llegar a esta adquisición son necesarios toda una serie de estudios previos que están detallados por diversos autores. Tanto si se ha realizado un estudio de mercado previo como un estudio piloto, en definitiva se necesitará una solicitud formal de ofertas a los proveedores de sistemas del mercado. Dicha solicitud habrá de especificar los requisitos contractuales y la metodología de evaluación.

Los requisitos contractuales que deben ser incluidos son la aceptación de soluciones con varios proveedores, provisión de desarrollo especial del software, la adecuación de los elementos preexistentes, las condiciones generales de la propuesta y un borrador de las condiciones del contrato.

La solución óptima a los complejos requisitos de un SIG puede implicar una variedad de proveedores de hardware v software. La empresa ha de especificar si quiere un suministrador principal que coordine todo el proceso, aceptando la responsabilidad, o varios diferentes que pueden ser contratados para implementar las diversas partes del sistema bajo la dirección de la plantilla de la empresa. El suministrador principal puede ser tanto un único proveedor como una compañía especializada en integración de sistemas. La solicitud de ofertas ha de evaluar de igual modo si se ha de dirigir el proceso en su totalidad o si se pueden considerar propuestas para partes concretas.

La empresa debe especificar también si las soluciones propuestas se deben ajustar a un modelo ya cerrado o actualmente en vías de desarrollo. Hay riesgos en ambas posiciones,



ya que los sistemas en desarrollo pueden prometer mayores beneficios que otros, pero hay un riesgo significativo de que se desarrollen sus especificaciones en cuanto a ejecución o tiempo. El riesgo de eliminar estos sistemas del proceso radica en renunciar a los beneficios potenciales a mediano y largo plazo.

Una aproximación al problema es realizar una prueba de control (*benchmark test*) y fijar la fecha del mismo como límite. Sólo aquellos que para entonces tengan los requerimientos demostrables serán considerados.

El desarrollo especial del software debe incluir la personalización de los interfaces de usuario, el software de traducción de los datos existentes, funciones de procesamiento únicas, e interfaces con otros sistemas. Se debe especificar así mismo el proceso de desarrollo y actualización de las especificaciones del software y del diseño, y el seguimiento de la implementación.

Las condiciones generales de la oferta deben incluir la fecha límite, la información mínima para la oferta formal, las condiciones de variación de propuestas durante el tiempo de evaluación y los precios base. La propuesta exige a los proveedores responder a un gran número de requisitos técnicos y contractuales. De la misma manera pueden ser requeridos para explicar cómo sus propuestas se adecúan a las funciones y ejecuciones obligatorias, deseables y opcionales. Deben adaptarse a las condiciones genéricas del sistema y a las condiciones del borrador del contrato. La empresa puede incluir un cuestionario para reflejar más claramente los aspectos que más interesen para evitar respuestas vagas.



La metodología de evaluación sería un índice de los procesos de preselección, una prueba de control de evaluación de ofertas y evaluación de costo-efectividad, y una descripción general de los criterios de evaluación de cada paso.

Esta solicitud de ofertas es enviada finalmente a los proveedores por cualquiera de los medios existentes en la actualidad. Para permitir a los proveedores preparar convenientemente sus propuestas, es recomendable contemplar un periodo de al menos ocho semanas. Durante este periodo se puede proponer una reunión formal con ellos.

El siguiente paso dentro de la adquisición de tecnología sería la propia evaluación de alternativas. Para determinar cuál es la más adecuada. Boehm (1981) describe varios modelos de costo/efectividad y de actuación que proporcionan criterios de decisión. Clarke proporciona uno basado en ellos.

En primer lugar se haría una preselección de ofertas, cuyo objetivo sería una lista breve de sistemas posibles mediante la evaluación y puntuación de la información suministrada por los proveedores. La evaluación previa de propuestas detalladas debe identificar cualquier relación entre las mismas, y si deben ser rechazadas sin evaluación posterior. Las causas del rechazo en primera instancia pueden ser la falta de adecuación con los requerimientos obligatorios, la falta de detalles, un desarrollo inaceptable del sistema, la imposibilidad de formar parte de una solución global, y unos costos que excedan considerablemente de las alternativas y del presupuesto del proyecto.

Las funcionalidades no obligatorias y los requisitos genéricos del sistema se puntúan a partir de las respuestas dadas por los proveedores. Se asigna un peso a cada requisito y se puntúa según una escala numérica (Boehm 1981). Los pesos se determinan mediante consulta a los usuarios, antes de recibir las ofertas. La experiencia ganada si se ha realizado un estudio piloto supondrá una base para determinar la importancia relativa de



las funciones. El diálogo con otros usuarios ayudará enormemente a la puntuación de aspectos como la calidad del mantenimiento y la asistencia técnica.

Como resultado de esta evaluación previa se tendrá una lista de ofertas, que se recomienda que no sea de más de cinco para que el proceso de control de las mismas no sea demasiado engorroso.

A continuación se realiza una prueba de control, cuyo objetivo es confirmar la puntuación de los requisitos funcionales y determinar estimaciones reales de su actuación en términos de capacidad de trabajo. Las partes son: el diseño del control, el desarrollo de los datos y la documentación de control, y el análisis de los resultados.

El diseño se debe basar en los requisitos especificados en la solicitud de ofertas. Debe establecer las tareas a desarrollar, los datos sobre los que se desarrollará y las salidas requeridas. La documentación debe proveer una descripción general de las tareas y una copia de los datos a utilizar. Los proveedores pueden prepararse para el control cargando los datos existentes y asegurándose con los conocimientos apropiados y la experiencia disponible, pero no es necesario ni apropiado especificar detalles de cada tarea que se desarrollará en el proceso.

Han de realizarse anotaciones con cuidado durante toda la ejecución del control. La configuración, la carga y la versión del software usado han de ser unas de ellas. Los evaluadores han de asegurarse que entienden lo que se les está demostrando y que todas las funciones se ejecutan en tiempo real.

Los resultados de la prueba de control permitirán refinar la puntuación de funcionalidad y la valoración de puntuaciones para la capacidad de trabajo. Las ofertas que no reúnan las



condiciones de funcionalidad obligatorias, o que no puedan proporcionar el mínimo de capacidad de trabajo, se eliminan en este momento.

Goodchild y Rizzo (1987) distinguen entre lo que ellos llaman control cualitativo y cuantitativo. El propósito de un control cualitativo es determinar el grado en el que el sistema propuesto puede realizar las funciones requeridas para satisfacer al equipo de control. Proponen una escala de "inhibición" para valorar el grado en que una función dada entra dentro de una actuación ideal, y así mismo inhibe la capacidad del sistema para generar productos particulares que dependan de la función. Se usa un control cuantitativo para valorar el grado en que el sistema puede verdaderamente realizar la capacidad de trabajo requerida sin los condicionantes del tiempo de trabajo del personal, la velocidad de la CPU, los dispositivos periféricos y la capacidad de almacenamiento.

Cada función requerida es probada al menos una vez durante la prueba. Su valoración cualitativa es dada por el equipo de control y se anota su utilización como recurso a través de varias medidas del tamaño del problema. Entonces se usan para construir modelos de predicción de utilización de recursos para cada función, de manera que la capacidad de trabajo se puede estimar dados los tamaños anticipados de los problemas de producción. El resultado es una serie de estimaciones del total de utilización de recursos, que pueden ser comparadas con las capacidades del sistema propuesto. En un ejemplo (Goodchild 1987), el control cuantitativo mostró que el proveedor había sobrestimado gravemente las tasas de digitalización que podía alcanzar la producción y también sobre configurado la CPU del sistema.

La última parte consiste en una evaluación de costo-efectividad. Las ofertas que sobrevivan a los dos procesos anteriores son evaluadas finalmente a través de su efectividad de costo. Se realizan una serie de actividades como la formación de



configuraciones teóricas, el análisis de los costos de cada configuración, el cálculo de las tasas de costo-efectividad y el análisis de los resultados.

Las configuraciones teóricas se crean definiendo el hardware y el software requeridos. Se puede requerir una normalización del hardware en cuanto al volumen de almacenamiento en disco y el número de estaciones de trabajo.

Los costos operativos de capital y recurrentes para cada configuración se determinan entonces para una vida nominal del proyecto de al menos cinco años. Mientras sólo se requieren realmente las diferencias de costo para evaluar las distintas alternativas de configuración, los costos totales deben ser también determinados para asegurar que el análisis costo-beneficio original sigue siendo válido. Las relaciones deben mostrar los costos de capital y operativos para cada año, tanto a precio constante como a valor actual.

La tasa de costo-efectividad para cada configuración se evalúa a continuación dividiendo el costo del valor actual para el conjunto de la vida nominal por la puntuación de la ejecución funcional y de la capacidad de trabajo, dando el costo por unidad teórica de ejecución. Dado que aquellos sistemas que no cumplan los niveles mínimos requeridos de funcionalidad y capacidad de trabajo no aparecerán en la lista final, las tasas son realmente una medida del incremento marginal de efectividad que se puede alcanzar con cada configuración superviviente.

Aunque la configuración con la mejor tasa de costo-efectividad es en teoría la elección óptima, puede darse el caso de que no sea realizable y puede haber otros factores no incluidos en la evaluación que deberían ser considerados también. Dentro de estos estarían las diferencias en el impacto que cada configuración tenga en la empresa y el



personal, y aspectos relativos a la viabilidad financiera de los proveedores. El informe final del proceso de evaluación incluirá, por consiguiente, para cada configuración de la lista final, los inventarios de costos totales, la puntuación para la funcionalidad no obligatoria, los requerimientos genéricos y de ejecución, un informe sobre los factores relevantes que no han sido incluidos en los costos y puntuaciones, y por último una revisión del análisis costo-beneficio original.

Al final de la selección sólo quedaría la adquisición propia del sistema. Una vez adquirido se pasaría a realizar todos los aspectos que se han tenido en cuenta hasta el momento, como son la construcción de las aplicaciones (programación, etc.) y la construcción de la base de datos (con sus aspectos de generación y conversión de datos, etc.).

Dentro de la empresa habrá que tratar para todo este proceso con diversos interlocutores, pasando desde el nivel de gestión, donde se contactará a los responsables económicos, hasta el nivel operativo, con los técnicos de sistemas.

Todos los procesos se reflejarán en los formularios convenientes, algunos de los cuales ya han sido indicados, como son la lista de preselección de ofertas, la propia evaluación de ofertas, el Análisis Costo-Efectividad, y las pruebas de funcionalidad, rendimiento e integridad.

Fruto de toda esta parte se tendrá una gran variedad de documentos. Estará toda la documentación de adquisición, y de la propia de construcción del sistema, como pueden ser los procesos de captura de datos, la programación de las aplicaciones, etc.



Implantación.

Es la última parte de la fase de implementación. En primer lugar habrá que dotar al sistema de infraestructura de soporte. La propia instalación del sistema implicará preparar locales, establecer el sistema de comunicaciones y el desarrollo de los interfaces propios y especiales del software. Se han diseñado toda una serie de pruebas para asegurar que las especificaciones contratadas en el sistema se pueden realmente obtener bajo condiciones normales.

Se necesitará una adaptación a los datos en explotación del sistema (captura inicial de datos y desarrollo de los productos), así como a los nuevos procedimientos de manipulación y actualización. La adaptación deberá llevarse a cabo también en las aplicaciones en explotación. La formación puede ser hecha en fases para apoyarse en la experiencia ganada. La efectividad de la formación se evaluará formalmente después de cada fase y los resultados revisados con el proveedor. La captura inicial de datos y los productos pueden ser también evaluados conjuntamente con los usuarios y, en caso de existir problemas, con el proveedor.

Una vez el sistema se encuentre en funcionamiento, es necesario un equipo de soporte que se encargue del mantenimiento de las aplicaciones y de las bases de datos. Su función es velar por que el sistema trabaje correctamente y reparar las posibles deficiencias, bien a través de su experiencia o con asistencia de los proveedores.

Dado el nivel físico al que se trabaja, los interlocutores en este caso serán, dentro del nivel de gestión, los jefes de área o servicio, y a nivel operativo, los usuarios finales que manejarán el sistema.



Los formularios deberán recoger todos estos aspectos de la implantación, como son la infraestructura de soporte y los nuevos procedimientos a aplicar.

De la misma manera se recogerán las actuaciones en documentos, en los que figurará la política de servicio, los nuevos procedimientos de trabajo finales, y toda la documentación de usuario (manuales, etc.).

Fase 6: Puesta en marcha.

Por último, se llega a la puesta en marcha definitiva. En esta fase se pretende evaluar los resultados que se obtienen al final del proceso. De esta puesta en marcha se verán nuevas necesidades dentro de la empresa, y se pasará a revisar todo el proceso de la metodología, produciéndose una retroalimentación de todo el conjunto.

Habrá que revisar la organización en producción, si está dando los resultados previstos, y si los nuevos procedimientos en funcionamiento se adecúan con la organización y el sistema. El impacto que tenga el SIG en la organización habrá de ser estudiado, ante actitudes de rechazo, tiempos, etc., que pueden dar resultados negativos. Por último, habrá que revisar la generación de productos de información, si estos son los previstos, y si se producen tal y como se pretendía al instalar el sistema. Estos productos pasarán en un principio a generar únicamente mapas, para con el tiempo y la experiencia desarrollar el campo del análisis y de la gestión integral, que es para lo que, en definitiva, se implementa el SIG.

Para hacer esta revisión, los interlocutores irán desde el nivel de decisión con los gerentes de la empresa, el nivel de gestión con los jefes de área o servicio, hasta el nivel operativo con los técnicos y usuarios finales.



En esta última fase los documentos que se obtienen son los productos de información para los que toda esta metodología de implantación del SIG se ha desarrollado, en la medida en que son el objetivo último de todo el proceso.

Para mayor información sobre las actividades de catastro de la OEA haga clic [aquí](http://www.oas.org/es/sap/dsmg/catastro/) o visítenos en <http://www.oas.org/es/sap/dsmg/catastro/>

Visítenos también en

