

Álgebra para la gestión de coberturas y su integración en XQuery

Diego Loureda Couceiro¹
José Ramón Ríos Viqueira²
José Varela Pet³

Pedro Álvarez Pérez-Aradros⁴

Universidad de Santiago de Compostela (España), loureda@usc.es¹, joserios@usc.es², eljpet@usc.es³
Universidad de Zaragoza (España), alvaper@unizar.es⁴

Resumen: En este artículo se propone un álgebra para la gestión de coberturas geográficas y su integración en el lenguaje XQuery de consulta de documentos XML. Se formaliza el concepto de cobertura geográfica como un conjunto de funciones, llamadas atributos o bandas. Cada banda de una cobertura asocia a cada localización de un espacio discreto, finito y bidimensional un valor de un determinado dominio convencional (reales, enteros, cadenas de caracteres, etc.). Las operaciones del álgebra propuesta permiten la combinación de las bandas de dos coberturas, y el cómputo de bandas derivadas de las bandas ya existentes en una cobertura. La incorporación de estas operaciones en XQuery, lo dotan de capacidades de análisis espacial, convirtiéndolo así en un lenguaje adecuado para la gestión de fuentes de datos en Geography Markup Language (GML). Estas operaciones se ilustran mediante ejemplos representativos de su funcionalidad y mediante un ejemplo de aplicación realista.

INTRODUCCIÓN

Tradicionalmente, los datos geográficos se clasifican en dos grandes categorías: objetos y coberturas. Un objeto geográfico (una ciudad, una carretera o un municipio), además de poder poseer un conjunto de propiedades de tipo alfanumérico (población de la ciudad, gestor de la carretera o nombre del municipio), incluye una o varias propiedades de algún tipo geométrico, que definen su posición y forma en el espacio geográfico (superficie terrestre). Dichos tipos geométricos incluyen los bien conocidos puntos, líneas y superficies. Una cobertura geográfica asocia a cada punto de un determinado subconjunto del espacio geográfico, un conjunto de propiedades de tipo alfanumérico. El tipo de cada propiedad puede ser tanto discreto como continuo. Ejemplos de propiedades discretas son el tipo de suelo (planeamiento urbanístico) y tipo de vegetación. Ejemplos de propiedades continuas son la temperatura y la altitud sobre el nivel del mar.

Los primeros sistemas de información dedicados a la gestión de información geográfica (Sistemas de Información Geográfica, SIG), se diseñaban para ámbitos de aplicación específicos y almacenaban tanto las propiedades de los objetos geográficos como las coberturas en el sistema de archivos del sistema operativo. Por lo tanto, las bien conocidas ventajas del uso de Sistemas Gestores de Bases de Datos (SGBD) no se aprovechaban en estos sistemas. Una primera evolución de estos sistemas se produce con el desarrollo de herramientas SIG de propósito general. Estas primeras herramientas mantenían el uso de archivos, y la gestión de datos geográficos se realizaba mediante la ejecución de grandes listas de comandos, que en el caso de la gestión de coberturas estaban basados en el álgebra de Tomlin [8]. Si nos restringimos a la gestión de objetos geográficos, las herramientas SIG pronto empezaron a utilizar un SGBD convencional para almacenar las propiedades alfanuméricas, y en algunos casos también las geométricas. Este SGBD convencional ya proporciona funcionalidad genérica de control de concurrencia, back-up, gestión de transacciones, seguridad, etc. Sin embargo, la ausencia de un lenguaje de consulta para datos geométricos hace que su gestión tenga que hacerse en una capa software fuera del gestor, disminuyendo así la eficiencia del sistema. Para solucionar estos problemas, se investiga en la extensión de los SGBD con funcionalidad de consulta espacial, alcanzando como resultado las actuales extensiones espaciales de los SGBDs más populares (Oracle, DB2, PostgreSQL, etc.). Es importante resaltar que no existe una solución satisfactoria para la gestión de coberturas geográficas desde dentro de un SGBD.

La gran importancia de Internet en la nueva sociedad de la información hace que el uso de Servicios Web en SIG sea actualmente la tendencia dominante. En este sentido, el Open Geospatial Consortium (OGC) proporciona especificaciones estándar para Servicios Web de acceso a datos. Ejemplos reseñables son el Web Feature Service (WFS), para acceso a datos de objetos geográficos, el Web Coverage Service (WCS) para acceso a datos de coberturas geográficas, y el Geography Markup Language (GML), lenguaje XML de transferencia de datos geográficos (objetos y

coberturas) entre los servicios web. Ninguno de los dos servicios anteriores proporciona una capacidad de análisis espacial similar a la que proporciona un SGBD espacial para objetos geométricos.

En este artículo se presenta un álgebra que permite el análisis espacial de datos obtenidos de coberturas geográficas y se muestra la forma en la que este álgebra se integra en el lenguaje XQuery. Más en concreto, la contribución del presente trabajo puede resumirse en los siguientes puntos:

- Se formaliza el concepto de cobertura geográfica como un conjunto de funciones, llamadas atributos o bandas, que asocian a cada localización de un espacio discreto de dimensión 2 un valor de un dominio convencional (enteros, reales, cadenas de caracteres, etc.)
- Se da una descripción informal de las dos operaciones que permiten la combinación de las bandas de dos coberturas y el cómputo de nuevas bandas derivadas de las ya existentes. Estas operaciones se ilustran mediante algunos ejemplos representativos, que permiten alcanzar la funcionalidad de una operación de cada uno de los tipos descritos por Tomlin en [8].
- Se muestra también la forma en la que el álgebra propuesta se integra en el lenguaje de consulta sobre XML XQuery. Dicha integración permite el uso de XQuery para la consulta de fuentes de datos de coberturas geográficas en GML [6].

Los tres puntos anteriores marcan el contenido de las tres secciones siguientes de este artículo. Además, el artículo incluye dos secciones más, una en la que se presenta un ejemplo de aplicación realista para el álgebra propuesta y otra en la que se resumen las conclusiones y las líneas de trabajo futuro.

COBERTURAS GEOGRÁFICAS

Una *cobertura geográfica* se define como un conjunto de funciones, cada una de las cuales asocia un valor de un determinado dominio alfanumérico (enteros, reales, cadenas de caracteres, etc.) a cada punto de una determinada superficie del *espacio geográfico*. En esta sección se formaliza e ilustra mediante ejemplos el concepto de cobertura geográfica.

Espacio geográfico y localizaciones

Sean R y Z los conjuntos de los números reales y enteros, respectivamente, y sea n un número entero mayor que cero cualquiera. El conjunto de puntos del *espacio geográfico* se representa mediante el conjunto infinito

$$R^2_n \equiv \{(x, y) \mid x, y \in R \wedge -n-0.5 \leq x, y < n+0.5\}.$$

En el presente modelo, por lo tanto, el espacio geográfico es un rectángulo cerrado en sus lados inferior e izquierdo y abierto en sus lados superior y derecho. Si i, j son dos números enteros tal que $-n \leq i, j \leq n$, entonces una *localización* $z_{i,j}$ del *espacio geográfico* se define como el subconjunto de R^2_n

$$z_{i,j} \equiv \{(x, y) \mid x, y \in R \wedge i-0.5 \leq x < i+0.5 \wedge j-0.5 \leq y < j+0.5\}.$$

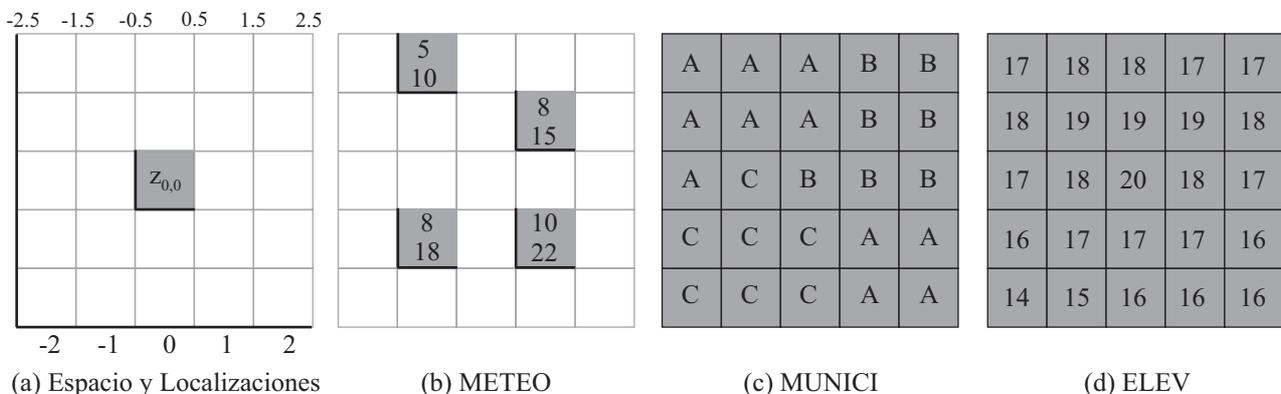


Figura 1: Espacio, localizaciones y coberturas geográficas

Una localización es por lo tanto un subconjunto infinito de puntos del espacio geográfico, con forma cuadrangular y con dos lados cerrados (izquierdo e inferior) y dos lados abiertos (superior y derecho). Se dice que (i, j) son las *coordenadas discretas* de la localización $z_{i,j}$. Se dice que el conjunto finito de localizaciones $Z_n^2 = \{z_{i,j} \mid i, j \in Z \wedge -n \leq i, j \leq n\}$ es una *representación discreta* (finita) del espacio geográfico. Claramente, Z_n^2 es una partición del espacio geográfico, por lo tanto, cada punto del espacio geográfico pertenece a una y sólo una localización de Z_n^2 . La Figura 1(a) ilustra los conceptos anteriores para $n = 2$.

Dominios y coberturas geográficas

Un *dominio geográfico* es un subconjunto S del *espacio geográfico* ($S \subseteq R_n^2$) que se puede expresar como la unión de un conjunto finito de localizaciones, es decir,

$$S = \bigcup_i p_i, p_i \in Z_n^2.$$

Sea S un *dominio geográfico*, A_1, A_2, \dots, A_m nombres distintos y D_1, D_2, \dots, D_m nombres de dominios convencionales, no necesariamente distintos. Una *cobertura geográfica* C con esquema $C[S](A_1 \mid D_1, A_2 \mid D_2, \dots, A_m \mid D_m)$ se define como un conjunto de funciones

$$A_i: \{p \in Z_n^2 \mid p \subseteq S\} \rightarrow D_i \cup \{\text{undefined}\}, i = 1, 2, \dots, m.$$

Cada función A_i se denomina *atributo* o *banda*, y asocia a cada localización p contenida en S un valor de un dominio convencional (real, entero, cadena de caracteres, etc.) o el valor especial *undefined* (A_i no está definida en p). En las Figuras 1(b), 1(c) y 1(d) se muestran las representaciones gráficas de tres coberturas. La primera, con esquema METEO[S₁](temperatura | real, humedad | real), asocia un valor de temperatura y otro de humedad, medidos por una estación meteorológica, a la localización de dicha estación. La segunda, MUNICI[S₂](nombre | string) asocia a cada localización de su dominio el nombre del municipio al que pertenece. La tercera, ELEV[S₂](elevación | real) almacena la elevación sobre el nivel del mar en cada localización de su dominio. En lo que queda de artículo, A y B posiblemente con un subíndice se utilizan para denotar bandas cuyo tipo es irrelevante.

ÁLGEBRA DE COBERTURAS GEOGRÁFICAS

En esta sección se muestran las operaciones de combinación de coberturas y de cómputo de bandas, que incluye la presente álgebra de gestión de coberturas geográficas. No se pretende la formalización de estas operaciones, sino la ilustración de su funcionalidad mediante ejemplos representativos. En concreto, se muestran ejemplos de como podrían expresarse operaciones entre coberturas de cada uno de los tipos propuestos en [8], que son la base de la mayoría de sistemas de gestión de coberturas comerciales [3, 4, 5] y de libre distribución [1, 7] disponibles en la actualidad.

Combinación de coberturas

Operación Combine: Sean C_1 y C_2 dos coberturas geográficas con esquemas respectivos, $C_1[S_1](A_1, A_2, \dots, A_n)$, $C_2[S_2](B_1, B_2, \dots, B_m)$, donde no existe el mismo nombre de banda en C_1 y C_2 . La operación

$$C = \text{Combine}(C_1, C_2)$$

devuelve una nueva cobertura C , con esquema $C[S_1 \cup S_2](A_1, A_2, \dots, A_n, B_1, B_2, \dots, B_m)$, donde para cada localización $p \subseteq S_1 \cup S_2$

- Si $p \subseteq S_1 \cap S_2$ entonces $C.A_i(p) \equiv C_1.A_i(p)$, $\forall i = 1, 2, \dots, n$ y $C.B_j(p) \equiv C_2.B_j(p)$, $\forall j = 1, 2, \dots, m$.
- Si $p \subseteq S_1 - S_2$ entonces $C.A_i(p) \equiv C_1.A_i(p)$, $\forall i = 1, 2, \dots, n$ y $C.B_j(p) \equiv \text{undefined}$, $\forall j = 1, 2, \dots, m$.
- Si $p \subseteq S_2 - S_1$ entonces $C.A_i(p) \equiv \text{undefined}$, $\forall i = 1, 2, \dots, n$ y $C.B_j(p) \equiv C_2.B_j(p)$, $\forall j = 1, 2, \dots, m$.

Claramente, la cobertura resultado C tendrá los valores originales de C_1 y C_2 para sus bandas A_i y B_j en las localizaciones en las que tanto C_1 como C_2 están definidas, es decir, en $S_1 \cap S_2$. En las localizaciones donde C_1 está definida pero C_2 no lo está, las bandas A_i tendrán el valor original de C_1 y las bandas B_j tendrán valor nulo. Finalmente, en las localizaciones donde C_1 no está definida pero C_2 sí lo está, las bandas A_i tendrán valor nulo y las bandas B_j tendrán el valor original de C_2 .

Cómputo de bandas

Para obtener nuevas bandas a partir de las ya existentes en una cobertura, se utilizan *sentencias de cómputo*. No es objetivo de esta subsección el dar una definición exhaustiva de todas las posibles sentencias de cómputo soportadas en la presente álgebra, sino la ilustración mediante ejemplos representativos de la funcionalidad que puede alcanzarse.

Data una cobertura geográfica con esquema $C_1[S_1](A_1, A_2, \dots, A_n)$, una *sentencia de cómputo* tiene la siguiente estructura:

```
FOR EACH <loc1>, <loc2>
WHERE <condition>
COMPUTE B1 = e1, B2 = e2, ..., Bm = em
```

$\langle loc_1 \rangle, \langle loc_2 \rangle$ ($\langle loc_2 \rangle$ es opcional) son nombres de variables que iteran de forma independiente sobre el conjunto de localizaciones contenidas en S_1 . La cláusula “**WHERE** $\langle condition \rangle$ ” es opcional y se evalúa para cada combinación de las anteriores variables, seleccionando sólo aquellas combinaciones de localizaciones para las que $\langle condition \rangle$ devuelve un valor verdadero. Finalmente, el resultado de la evaluación de cada expresión e_j ($j=1, 2, \dots, m$) para cada combinación de localizaciones obtenidas del paso anterior, será el valor asociado a la localización referenciada por $\langle loc_1 \rangle$ en la nueva banda B_j ($j=1, 2, \dots, m$) de la cobertura resultado.

Teniendo en cuenta la descripción anterior de las *sentencias de cómputo*, la siguiente operación permite la evaluación de una sentencia de cómputo sobre una cobertura.

Operación Evaluate: Sea C_1 una cobertura geográfica con esquema $C_1[S_1](A_1, A_2, \dots, A_n)$ y *fwc* una *sentencia de cómputo* válida. La operación

$$C = Evaluate[fwc](C_1)$$

devuelve una nueva cobertura C resultado de la evaluación de la sentencia *fwc* sobre la cobertura C_1 .

Ilustración de la funcionalidad alcanzada

En esta subsección muestra como puede expresarse con el álgebra propuesta, una operación entre coberturas de cada uno de los tipos propuestos en [8].

| | | | | |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 |
| 0 | 5 | 0 | 6 | 0 |
| 0 | 0 | 0 | 0 | 0 |
| 0 | 3 | 4 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 |

(a) EDIFICIOS

| | | | | |
|----|----|----|----|----|
| 17 | 18 | 18 | 17 | 17 |
| 18 | 24 | 19 | 25 | 18 |
| 17 | 18 | 20 | 18 | 17 |
| 16 | 20 | 21 | 17 | 16 |
| 14 | 15 | 16 | 16 | 16 |

(b) ELEV_TOTAL

Figura 2: Ilustración de la suma local de elevaciones (Figura 1(d)) y alturas de edificios.

Operaciones locales: Dadas un conjunto de bandas de entrada, la banda resultante de una operación local asigna a cada localización p un valor que se obtiene a partir de los valores asociados a esa misma localización p en las bandas de entrada. Así por ejemplo, si *ELEV* es la cobertura de elevaciones del terreno de la Figura 1(d) y *EDIFICIOS*[S_2](*elev_edif* | *real*) es la cobertura de elevaciones de edificios mostrada en la Figura 2(a), la elevación total en cada localización se obtiene mediante la suma de las dos anteriores. Este es el resultado que devuelve la siguiente secuencia de operaciones del álgebra propuesta.

$$C = Combine(ELEV, EDIFICIOS)$$

$$ELEV_TOTAL = Evaluate[FOR EACH p COMPUTE elev_total(p) = elevacion(p) + elev_edif(p)](C)$$

El esquema de la cobertura resultante es $ELEV_TOTAL[S_2](elev_total | real)$, y su representación gráfica se muestra en la Figura 2(b). Es importante recordar en este punto que tanto *elevacion* como *elev_edif* son funciones que aplicadas a una localización devuelve un valor real. Para cada localización p de C el valor de $elev_total(p)$ se obtiene de la evaluación de la expresión $elevacion(p) + elev_edif(p)$. Las referencias explícitas a la primera variable de una sentencia de cómputo (p en este ejemplo) se pueden eliminar de todas las bandas. Así, la expresión anterior podría escribirse también como: $elev_total = elevacion + elev_edif$. Esta decisión de diseño tiene como objeto simplificar la sintaxis de las sentencias. Más adelante veremos que estas referencias no podrán eliminarse para las variables que no son la primera ni para funciones que no sean bandas.

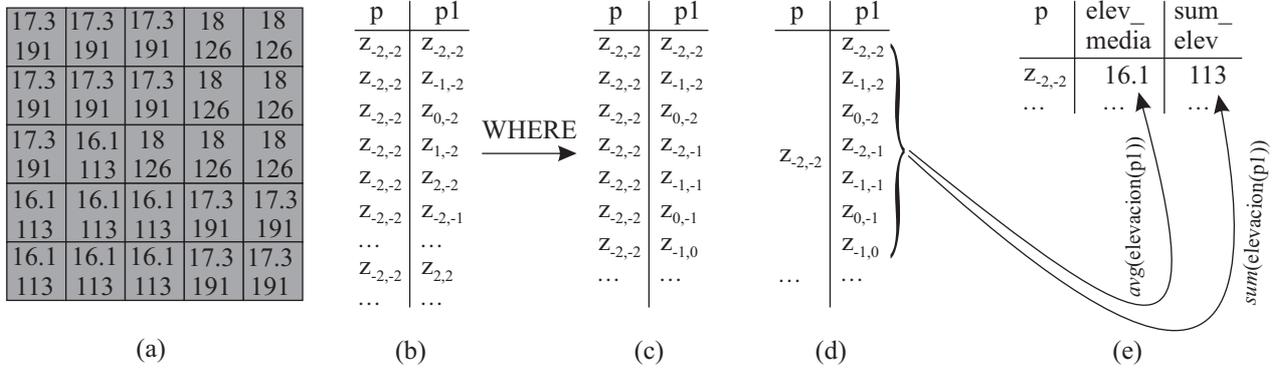


Figura 3: Ilustración de media y suma zonales de elevaciones (Figura 1(d) en cada municipio (Figura 1(c)).

Operaciones zonales: Una zona en una banda de una cobertura la componen todas las localizaciones que tienen asignado el mismo valor. Teniendo esto en cuenta, dadas dos bandas de entrada A y B, una operación zonal devuelve para cada localización p un valor que se obtiene a partir de los valores que asigna la banda A a todas las localizaciones que están en la zona de p en la banda B. Por ejemplo, si MUNICI y ELEV son las coberturas mostradas en las Figuras 1(c) y 1(d), operaciones zonales podrían obtener para cada localización p la media y la suma de las elevaciones del municipio que contiene a p . Este resultado se obtiene mediante las operaciones siguientes.

$C = Combine(MUNICI, ELEV)$

$ELEV_MUNICIPAL = Evaluate[FOR EACH p, p1$
WHERE nombre = nombre(p1)
COMPUTE elev_media = avg(elevacion(p1)),
 sum_elev = sum(elevacion(p1))](C)

El esquema de la cobertura resultado es $ELEV_MUNICIPAL [S_2](elev_media | real, sum_elev | real)$ y su representación gráfica se muestra en la Figura 3(a). Para obtener este resultado, en primer lugar se obtienen todas las combinaciones ($p, p1$) de localizaciones en el dominio de C (ver Figura 3(b) para $p = z_{-2,-2}$). Después la cláusula WHERE restringe esas combinaciones a aquellas en las que el valor de la banda *nombre* en p es igual al valor de la banda *nombre* en $p1$, es decir, $nombre(p) = nombre(p1)$ (ver Figura 3(c) para $p = z_{-2,-2}$). En este caso la referencia a la variable $p1$ no puede eliminarse de la expresión. Implícitamente, antes del cómputo de las expresiones, las combinaciones ($p, p1$) resultantes se agrupan por p . Así, para cada grupo, es decir, para cada localización p , se obtiene la colección de localizaciones $p1$ que están en su misma zona para la banda *nombre* (ver Figura 3(d) para $p = z_{-2,-2}$). Finalmente, el valor de la banda *elev_media* para cada p , se obtiene aplicando la función de agregado *avg* sobre las elevaciones en cada $p1$ de su grupo. De forma similar, la función de agregado *sum* permite calcular la suma de todas las elevaciones en cada grupo, es decir, en cada p (ver Figura 3(e) para $p = z_{-2,-2}$).

Operaciones focales: Dada una banda de entrada, el valor de la banda resultante de una operación focal para una localización p se obtiene de los valores de la banda de entrada en un vecindario de p . Este vecindario puede estar compuesto por ejemplo por todas las localizaciones situadas a una distancia de p menor que un determinado umbral. Un ejemplo típico de operación focal es la interpolación. Si METEO y MUNICI son las coberturas mostradas en las Figuras 1(b) y 1(c) respectivamente, el método de interpolación IDW (Inverse Distance Weighted) obtiene el valor de temperatura en cada localización de cada municipio a partir de los valores de las localizaciones vecinas, situadas a menos de d unidades de distancia. En concreto, para cada localización p de cada municipio, el valor T de la temperatura en p se obtiene mediante la formula

$$T = (\sum_i t_i/d_i) / (\sum_i 1/d_i), \quad d_i < d$$

donde t_i es el valor de temperatura que asocia la cobertura METEO a la localización p_i , y d_i es la distancia entre p y p_i . Las siguientes operaciones aplican el método IDW a las coberturas METEO y MUNICI, para una distancia $d = 2$.

$C = \text{Combine}(\text{METEO}, \text{MUNICI})$

$\text{METEO2} = \text{Evaluate}[\text{FOR EACH } p, p1$
WHERE $\text{distance}(p, p1) < 2$ and $\text{temperatura}(p1)$ IS DEFINED
 and $\text{humedad}(p1)$ IS DEFINED
COMPUTE
 $\text{temperatura} = \text{sum}(\text{temperatura}(p1)/\text{distance}(p, p1)) / \text{sum}(1/\text{distance}(p, p1))$,
 $\text{humedad} = \text{sum}(\text{humedad}(p1)/\text{distance}(p, p1)) / \text{sum}(1/\text{distance}(p, p1))](C1)$

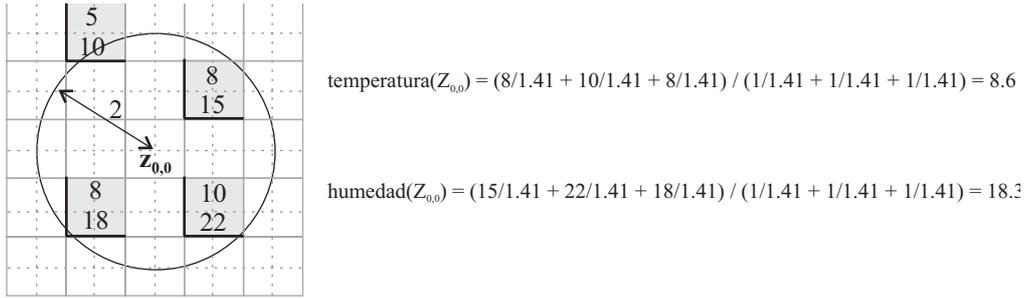


Figura 4: Ilustración de la interpolación IDW de la temperatura y humedad (Figura 1(b)) para la localización $z_{0,0}$.

Para cada localización p , la condición de la cláusula WHERE obtiene todas las localizaciones $p1$ situadas a una distancia menor que 2 y cuyo valor de temperatura está definido. Para esto se utiliza una función espacial distance , que calcula la distancia Euclidiana entre dos localizaciones, y el predicado IS DEFINED , que devuelve el valor booleano verdadero cuando se aplica a un valor distinto del valor especial undefined . La expresión de la cláusula COMPUTE aplica la fórmula del método IDW sobre el conjunto de localizaciones $p1$ obtenido tras la evaluación de la cláusula WHERE. La Figura 4 ilustra el cálculo de las bandas temperatura y humedad de la cobertura METEO2 para $p = z_{0,0}$.

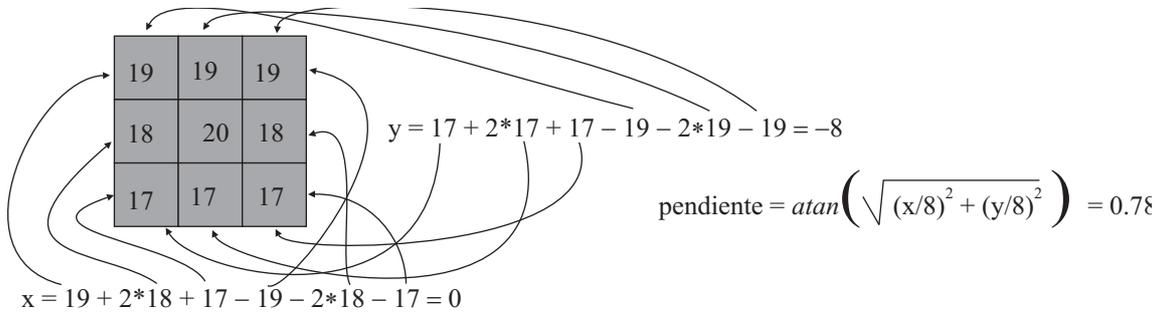


Figura 5: Ilustración del cómputo de la pendiente a partir de la elevación (Figura 1(d)) para la localización $z_{0,0}$.

Operaciones incrementales: Dada una banda de entrada, el resultado de una operación incremental asigna a cada localización p un valor obtenido a partir de los valores de las localizaciones adyacentes a p . Consideremos por ejemplo la cobertura de elevaciones del terreno ELEV de la Figura 1(d). La siguiente secuencia de operaciones obtiene una aproximación de la pendiente en cada localización, siguiendo la aproximación propuesta en [2].

$C1 = \text{Evaluate}[\text{FOR EACH } p$
WHERE $\text{elevation}(\text{north}(\text{west}(p)))$ IS DEFINED and $\text{elevation}(\text{north}(p))$ IS DEFINED
 and $\text{elevation}(\text{north}(\text{east}(p)))$ IS DEFINED and $\text{elevation}(\text{east}(p))$ IS DEFINED
 and $\text{elevation}(\text{south}(\text{west}(p)))$ IS DEFINED and $\text{elevation}(\text{west}(p))$ IS DEFINED
 and $\text{elevation}(\text{south}(\text{east}(p)))$ IS DEFINED and $\text{elevation}(\text{south}(p))$ IS DEFINED
COMPUTE $x = \text{elevation}(\text{north}(\text{west}(p))) + 2*\text{elevation}(\text{west}(p)) + \text{elevation}(\text{south}(\text{west}(p))) -$
 $\text{elevation}(\text{north}(\text{east}(p))) - 2*\text{elevation}(\text{east}(p)) - \text{elevation}(\text{south}(\text{east}(p)))$,
 $y = \text{elevation}(\text{south}(\text{west}(p))) + 2*\text{elevation}(\text{south}(p)) + \text{elevation}(\text{south}(\text{east}(p))) -$
 $\text{elevation}(\text{north}(\text{west}(p))) - 2*\text{elevation}(\text{north}(p)) - \text{elevation}(\text{north}(\text{east}(p)))$
 $](\text{ELEV})$

$$\text{PENDIENTE} = \text{Evaluate}[\text{FOR EACH } p \text{ COMPUTE pendiente} = \text{atan}(\text{sqrt}((x / 8)^2 + (y / 8)^2))](C1)$$

Dada una localización p con coordenadas (i, j) , las funciones espaciales $north(p)$, $south(p)$, $east(p)$ y $west(p)$, devuelven, respectivamente, las localizaciones con coordenadas $(i, j+1)$, $(i, j-1)$, $(i+1, j)$ y $(i-1, j)$. Las funciones convencionales sqrt y atan calculan, respectivamente la raíz cuadrada y el arcotangente de un número real. Finalmente, la operación r^e , devuelve el resultado de elevar el número real r a la potencia e . La Figura 5 ilustra el cómputo de la pendiente para $p = z_{0,0}$.

GESTIÓN DE COBERTURAS EN GML

En esta sección se ilustra mediante un ejemplo representativo la integración de las operaciones del álgebra de la sección anterior en el lenguaje de consulta XQuery, dotándolo así de funcionalidad de análisis espacial que lo habilita para la consulta de fuentes de datos de coberturas en GML.

Representación de coberturas en GML

En su última versión 3.1, el lenguaje GML [6] del OGC permite la representación de coberturas geográficas. GML soporta varios tipos de coberturas. Para el propósito del presente artículo, nos restringimos a sólo tres tipos: coberturas multipunto, coberturas multisuperficie y coberturas de grid rectificado. Una cobertura de tipo multipunto asocia a cada punto (representado con un par de coordenadas vectoriales) de su dominio geográfico, una tupla de valores de tipo convencional, un valor para cada banda de la cobertura. Así, por ejemplo, el siguiente trozo de código representa la cobertura METEO de la Figura 1(b) mediante una cobertura multipunto de GML.

```
<METEO xmlns="http://labsis.usc.es/coberturas">
  <gml:domainSet>
    <gml:MultiPoint>
      <gml:pointMember><gml:Point><gml:pos>-1 -1</gml:pos></gml:Point></gml:pointMember>
      <gml:pointMember><gml:Point><gml:pos> 1 -1</gml:pos></gml:Point></gml:pointMember>
      <gml:pointMember><gml:Point><gml:pos> 1 1</gml:pos></gml:Point></gml:pointMember>
      <gml:pointMember><gml:Point><gml:pos>-1 2</gml:pos></gml:Point></gml:pointMember>
    </gml:MultiPoint>
  </gml:domainSet>
  <gml:rangeSet>
    <gml:DataBlock>
      <gml:rangeParameters><gml:CompositeValue><gml:valueComponents>
        <Temperatura uom="gradosC"> template </Temperature>
        <Humedad uom="porcentaje"> template </Humidity>
      </gml:valueComponents></gml:CompositeValue></gml:rangeParameters>
      <gml:tupleList>8 18 10 22 8 15 5 10</gml:tupleList>
    </gml:DataBlock>
  </gml:rangeSet>
</METEO>
```

El tag `<gml:domainSet>` se utiliza para describir el dominio geográfico de la cobertura, es decir, los puntos con coordenadas $(-1, -1)$, $(1, -1)$, $(1, 1)$ y $(-1, 2)$. El tag `<gml:rangeSet>` define las bandas de la cobertura, así como los valores convencionales asociados a cada punto del dominio. Así, en el ejemplo el punto $(-1, -1)$ tiene asociado el valor 8 de temperatura y el valor 18 de humedad. La cobertura MUNICI de la Figura 1(c) se podría representar mediante un GML similar al anterior, en este caso de tipo multisuperficie, reemplazando los cuatro puntos por cuatro polígonos vectoriales y reemplazando los valores de temperatura y humedad por una lista de nombres de municipios. La cobertura ELEV de la Figura 1(d) se puede representar mediante la siguiente cobertura GML de tipo grid rectificado.

```
<ELEV xmlns="http://labsis.usc.es/coberturas">
  <gml:rectifiedGridDomain>
    <gml:RectifiedGrid dimension="2">
      <gml:limits>
        <gml:GridEnvelope>
          <gml:low>0 0</gml:low><gml:high>4 4</gml:high>
        </gml:GridEnvelope>
      </gml:limits>
      <gml:axisName>x</gml:axisName><gml:axisName>y</gml:axisName>
      <gml:origin><gml:Point><gml:pos>-2 -2</gml:pos></gml:Point></gml:origin>
      <gml:offsetVector>1 0</gml:offsetVector>
      <gml:offsetVector>0 1</gml:offsetVector>
    </gml:RectifiedGrid>
  </gml:rectifiedGridDomain>
```

```

<gml:rangeSet>
  <gml:DataBlock>
    <gml:rangeParameters><gml:CompositeValue><gml:valueComponents>
      <Elevacion uom="metros"> template </Elevacion>
    </gml:valueComponents></gml:CompositeValue></gml:rangeParameters>
    <gml:tupleList>
      14 15 16 16 16
      16 17 17 17 16
      17 18 20 18 17
      18 19 19 19 18
      17 18 18 17 17
    </gml:tupleList>
  </gml:DataBlock>
</gml:rangeSet>
</ELEV>

```

Como puede verse en el ejemplo, el dominio se representa ahora mediante un tag *<gml:rectifiedGridDomain>*, en el que se especifica el origen del grid (punto con coordenadas (-2, -2)), dos vectores de desplazamiento *<gml:offsetVector>* con valores (1, 0) y (0, 1) que proporcionan una resolución y una orientación, y un número de celdas raster *<gml:limits>*.

Integración del álgebra de coberturas en XQuery

El siguiente trozo de código XQuery, aplica el método de interpolación IDW sobre la banda de temperatura de la cobertura METEO de la sección anterior.

```

LET $meteo := doc('meteo.gml')
LET $cadena := string-join(("FOR EACH p, p1",
  "WHERE distance(p, p1) < 2 and temperatura(p1) IS DEFINED",
  "COMPUTE",
  "temperatura = sum(temperatura(p1)/distance(p, p1))",
  "/ sum(1/distance(p, p1))"), " ")
LET $temperatura := Evaluate($cadena, $meteo)
RETURN $temperatura

```

En primer lugar, la cobertura METEO se recupera del archivo 'meteo.gml'. Después en la variable \$cadena se almacena la cadena de caracteres de la sentencia de cómputo que define la operación de interpolación que se pretende aplicar (ver la sección de ilustración de funcionalidad avanzada para más detalles sobre esta expresión). A continuación, se utiliza la función Evaluate para aplicar la sentencia de cómputo sobre la cobertura \$meteo, obteniendo como resultado una nueva cobertura, \$temperatura. La incorporación de esta función **Evaluate** permite la aplicación de la respectiva operación del álgebra definida en este artículo sobre variables de tipo cobertura. Finalmente, la última instrucción devuelve la cobertura calculada. Además de la función **Evaluate**, se incluye en XQuery una función **Combine**, que permite aplicar la operación del mismo nombre sobre dos variables de tipo cobertura.

EJEMPLO DE APLICACIÓN

En esta sección se ilustra un ejemplo realista de aplicación del XQuery anterior. En concreto, se dispone de tres coberturas de entrada distintas. Una cobertura que almacena para cada localización una elevación sobre el nivel del mar (ver Figura 6). Una cobertura que asocia a la posición de cinco estaciones meteorológicas un valor medido de temperatura y otro valor de humedad. Las posiciones de estas estaciones se muestran en la Figura 7. Por último se dispone de una cobertura que asocia a cada localización un valor de un modelo de combustible, obtenido del tipo de vegetación (ver Figura 8). A partir de estas coberturas, se pretende estimar un índice de riesgo de propagación de incendios aplicando las operaciones definidas en este artículo.

En primer lugar, se calcula la pendiente (ver Figura 9) a partir de la elevación. Este cómputo ya se ilustró en una sección anterior. En segundo lugar se calculan dos coberturas, una de temperatura (Figura 10) y otra de humedad (Figura 11) a partir de la cobertura de estaciones meteorológicas de la Figura 7. Para esto se utiliza el método de interpolación IDW ya descrito en secciones anteriores.

La cobertura que asocia un valor del índice de propagación de incendios a cada localización se obtiene mediante una suma ponderada de las coberturas pendiente, modelos de combustible, humedad y temperatura. Esta suma ponderada no es más que una operación local en el álgebra descrita en este artículo



Figura 6: Elevación sobre el nivel del mar

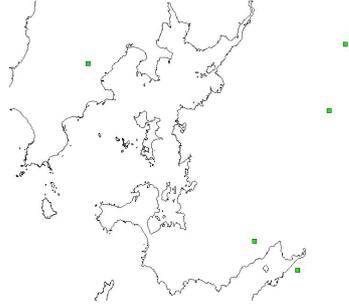


Figura 7: Estaciones meteorológicas

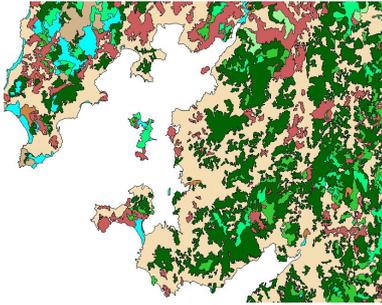


Figura 8: Modelos de combustible

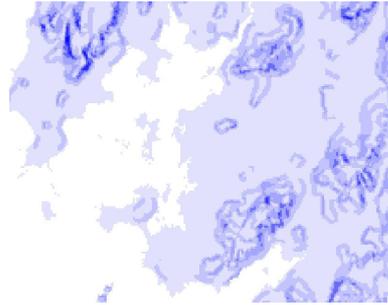


Figura 9: Pendiente

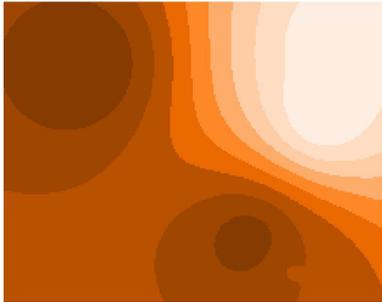


Figura 10: Temperatura

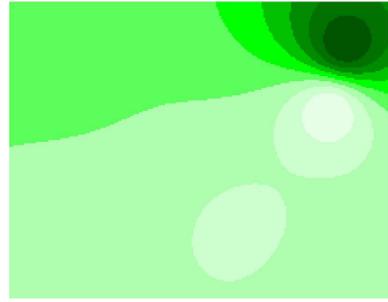


Figura 11: Humedad

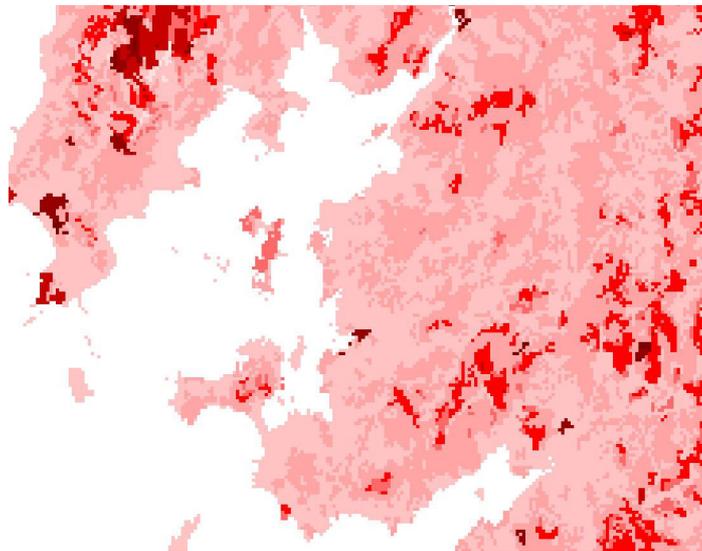


Figura 12: Riesgo de propagación de incendios

CONCLUSIONES Y TRABAJO FUTURO

En este artículo se ha ilustrado mediante ejemplos representativos la funcionalidad de las operaciones de un álgebra propuesta para la gestión de coberturas geográficas. La incorporación de estas operaciones al lenguaje de consulta XQuery, lo dotan de capacidades de análisis espacial, haciéndolo apropiado para la consulta de almacenes de datos GML de coberturas. Algunas de las ventajas más relevantes de la aproximación propuesta en este artículo, respecto a los sistemas actuales de gestión de coberturas geográficas son las siguientes.

- A diferencia de los sistemas disponibles en el mercado, las operaciones del álgebra propuesta son de carácter genérico. Así, por ejemplo, no se proporcionan dos operaciones distintas para el cálculo de la pendiente y el aspecto a partir de una cobertura de elevaciones, sino que se proporciona una herramienta (un lenguaje) que permite expresar estas operaciones y muchas otras más, mediante sentencias de cómputos de bandas.
- En el artículo se muestra una forma de integración del álgebra en XQuery. Sin embargo, la integración del álgebra en otros lenguajes de consulta también es posible. En concreto, la integración en el lenguaje SQL permitiría la gestión de coberturas desde dentro de un SGBD, característica no presente en ningún gestor comercial (a nuestro leal saber y entender).
- La combinación de las capacidades de consulta del lenguaje anfitrión (XQuery, SQL, etc.) con las capacidades de gestión de coberturas del álgebra propuesta permiten incrementar aún más la funcionalidad alcanzada. Así, por ejemplo, sería posible la utilización de operaciones del álgebra dentro de funciones recursivas de XQuery o dentro de consultas recursivas del nuevo SQL:2003.
- Las *sentencias de cómputo de bandas* proporcionan un lenguaje declarativo que permite expresar operaciones de cómputo sin la necesidad de programar la forma eficiente en que esas operaciones deberían ser implementadas. Esto permite, al igual que ocurre con los SGBD, la separación entre un nivel lógico en el que el usuario trabaja con conceptos más abstractos y sencillos (coberturas y sentencias de cómputo de bandas), y un nivel físico en el que el sistema utiliza complejas estructuras de datos y algoritmos para conseguir una implementación eficiente. Se permite por lo tanto al sistema la optimización de las sentencias expresadas por el usuario. Esta separación entre nivel lógico y físico, permite también al usuario el trabajar con coberturas discretas implementadas con un modelo vectorial (por ejemplo, las coberturas METEO y MUNICI de la Figura 1) y coberturas continuas implementadas con un modelo raster (por ejemplo la cobertura ELEV de la Figura 1) de una forma uniforme, sin preocuparse en ningún momento de qué modelo de representación interna (vectorial o raster) se utiliza para cada cobertura.
- La definición de nuevas operaciones entre coberturas geográficas y objetos geográficos, y la combinación con operaciones de álgebras de objetos geográficos ya existentes, permite la gestión uniforme de todo tipo de datos geográficos en un álgebra combinada de tipo heterogéneo (many-sorted).

Para terminar, las líneas de trabajo futuro incluyen:

- La formalización de la semántica de las sentencias de cómputo de bandas en un entorno formal de cálculo de predicados, así como la definición completa de la sintaxis del lenguaje.
- La implementación de un parser para el lenguaje anterior y el desarrollo de los algoritmos necesarios.
- La investigación en técnicas de optimización de sentencias de cómputo.
- La incorporación del lenguaje en un servicio de análisis de datos espaciales [9] que permita la consulta uniforme de fuentes de datos OGC, es decir, WFS y WCS.

REFERENCIAS

1. Geographic Resources Analysis Support System Homepage. (2005). <http://grass.baylor.edu/>. (último acceso: Noviembre 2005)
2. Horn, B.K.P. (1981). Hill Shading and the Reflectance Map, Proceedings of the IEEE 69(1):14-47.
3. Keigan Systems. (2005) MFWorks On Line. <http://www.keigansystems.com/software/mfworks/index.html> (último acceso: Noviembre 2005).
4. Lorup, E.J. (2000) IDRISI Tutorial online. <http://www.sbg.ac.at/geo/idrisi/wwwtutor/tuthome.htm> (último acceso: Noviembre 2005)
5. McCoy, J. and Johnston, K. (2001) Using ArcGis Spatial Analyst. Environmental Systems Research Institute, CA.
6. OGC-GML, 2004. OpenGIS Geographic Markup Language (GML) Encoding Specification. Open Geospatial Consortium. OGC 03-105r1. Version 3.1.1.
7. Red Hen Systems Inc. (2001) MapCalc User's Guide.
8. Tomlin CD (1990) Geographic Information Systems and Cartographic Modeling, Prentice Hall, Englewood Cliffs, NJ.
9. Viqueira J.R.R., Álvarez, P., Varela, J., Saco, P. (2005) Architecture of a natural disasters management framework and its application to risk assesment. Proc. 8th Conference on Geographic Information Science (AGILE 2005), 26-28 Mayo, Estoril, Portugal.