

Criterios para identificar reglas de asociación espacial entre objetos hidrológicos y coberturas terrestres en imágenes de satélite

Criteria for identification of spatial association rules between hydrological objects

Leisy Sosa Castro¹

Resumen El estudio integral del medio geográfico mediante el uso de información geoespacial en la actualidad constituye un reto tecnológico debido al gran volumen de información de diversas temáticas que contienen las imágenes de satélites. Ellas son portadoras de un gran número de objetos: vegetación, suelos, construcciones, litología, relieve y los objetos hidrológicos, objeto de estudio de esta investigación. Para desarrollar la investigación, primeramente se obtendrán los elementos que caracterizan el estado del arte de la temática, mediante el estudio de las fuentes bibliográficas existentes a escala nacional e internacional, para luego caracterizar los métodos y técnicas. Con tales antecedentes, el objetivo principal de la investigación es diseñar un método para extraer con la mayor eficiencia las asociaciones espaciales implícitas entre objetos hidrológicos y las coberturas terrestres en imágenes de satélites, así como su visualización y traducción a un lenguaje común, para favorecer el análisis ulterior en contextos más amplios en otras imágenes satelitales, y la identificación de objetos que permitan la toma de decisiones sobre el estado del medio ambiente geográfico y sus elementos componentes. Aspectos importantes en el orden científico son la caracterización de la taxonomía hidrológica a utilizar, la cual constituye el origen para el descubrimiento de las relaciones espaciales y no espaciales con el resto de los objetos ambientales que los circundan, así como la realización de los experimentos apropiados con el fin de diseñar y validar los algoritmos que darán soporte a la identificación y extracción de las reglas de asociación espacial, implícitas en las imágenes de satélites, así como su visualización y traducción a un lenguaje comprensible por los usuarios.

Abstract The comprehensive study of the geographical environment by using geospatial information currently constitutes a technological challenge due to the large volume of information on various topics containing satellite images. They are carrying a large number of items: vegetation, soil, construction, lithology, relief and hydrological objects, object of study of this research. To develop research, first the elements that characterize the state of the art theme, by studying existing literature sources at national and international level, and then characterize the methods and techniques will be obtained. With this background, the main objective of the research is to design a method to extract as efficiently implied hydrological spatial associations between objects and land cover in satellite imagery and visualization and translation into a common language to promote Further analysis in broader contexts other satellite images and identifying objects that allow decisions about the status of geographical environment and its components. Important in the scientific order aspects are the characterization of hydrological taxonomy to be used, which is the source for the discovery of spatial and non-spatial relationships with other environmental objects that surround them, and the undertaking of appropriate experiments in order to design and validate algorithms that will support the identification and extraction of spatial association rules implicit in satellite imagery and visualization and translation into a language understood by users.

Palabras Clave

Imágenes de satélites, objetos hidrológicos, coberturas terrestres, reglas de asociación espacial, relaciones implícitas

¹ Universidad Agraria de La Habana "Fructuoso Rodríguez Pérez", Facultad de Ciencias Técnicas, Cuba, lscastro@unah.edu.cu

1. Introducción

En la actualidad el manejo de grandes volúmenes de datos espaciales en imágenes de satélites trae consigo la demanda de nuevas técnicas para la extracción de relaciones u otros patrones no explícitos contenidos en dichas imágenes. Estas relaciones son muy útiles para encontrar e identificar objetos similares en contextos espaciales más amplios, es decir, en otras imágenes de satélites de territorios diferentes, de cara a una caracterización más eficiente de los objetos estudiados, que faciliten una evaluación más objetiva desde el punto de vista medioambiental y la toma de decisiones sobre el uso y cobertura de la tierra. Las técnicas tradicionales de análisis de datos no han tenido un desarrollo equivalente a su almacenamiento, pues el volumen de almacenamiento supera la capacidad de análisis [14].

En nuestro país, existen diversos programas medioambientales; entre ellos el Programa Nacional de Cambios Globales que aglutina más de una treintena de temáticas. La integración de sus resultados implica movilizar un numeroso grupo de expertos, no lográndose con técnicas tradicionales de análisis establecer las asociaciones entre dichas temáticas. Según Rosette [14] instituciones como el Ministerio de la Agricultura (MINAGRI), Ministerio de Ciencia Tecnología y Medio Ambiente (CITMA), Instituto Nacional de Recursos Hidráulicos (INRH) y el Grupo Empresarial GEOCUBA, entre otras, generan un gran volumen de datos geográficos que resultan importantes para la toma de decisiones en el ámbito económico y social.

Por otra parte, en Cuba no abundan los estudios para la extracción de relaciones implícitas entre los objetos hidrológicos y las coberturas terrestres de su entorno en imágenes de satélites. Tampoco se conocen algoritmos para la visualización de las reglas de asociación espacial entre los objetos hidrológicos y las clases de coberturas terrestres del entorno de las cuencas hidrográficas en imágenes de satélites, así como para la traducción de las reglas de asociación espacial a un lenguaje común más comprensible por el usuario del sistema.

De aquí que se hace necesario realizar una correcta abstracción del mundo real para el proceso de filtrado de los datos con el fin de garantizar una correcta valoración en la organización de las temáticas en la base de datos geoespacial, ya que los algoritmos tampoco ofrecen las condiciones de selección de los objetos geográficos que intervienen en el análisis para cada nivel de la taxonomía hidrológica diseñada. En este sentido la minería de datos espaciales como rama de la ciencia que se encarga de "extraer conocimiento referente a la naturaleza espacial de los datos" (Han y Kamber, [7]) constituye una derivación especial dentro de la minería de datos, lo que ayudaría notablemente al descubrimiento de estas asociaciones que permanecen subyacentes e invisibles a primera vista en las imágenes de satélites. La diferencia entre ambas categorías radica en que no sólo se emplean atributos no espaciales, sino que en el proceso de "minado" de los datos intervienen atributos espaciales combinados con los mismos, incorporándose también otros criterios como los predicados

espaciales. Estos predicados son las relaciones espaciales (Ej. Topológicas, Distancia, etc.) entre la temática de referencia y el conjunto de temáticas relevantes [3] por lo que se adopta un criterio complejo para descartar combinaciones de los objetos hidrológicos en la evaluación de predicados espaciales en las imágenes de satélites, haciendo el proceso más eficaz. Un aspecto esencial que se ha considerado, es que no existe suficiente experiencia en el empleo de las herramientas de las reglas de asociación espacial disponibles en software como gv-SIG, para incluir nuevas funcionalidades como visualización y traducción.

Es por ello que el presente trabajo está dirigido a la extracción de relaciones espaciales entre objetos hidrológicos y los objetos ambientales que los circundan en el entorno de las cuencas hidrográficas, que son desconocidas por el usuario; pero que están implícitas en las imágenes de satélites. En los momentos actuales, de acuerdo con las informaciones suministradas por el Centro de Aplicaciones de Tecnologías de Avanzadas (CENATAV) y constatadas con el Grupo de Investigaciones de Agrofísicas (GIAF) de la Universidad Agraria de la Habana (UNAH), en el país no se dispone de un método para la extracción de reglas de asociación entre objetos hidrológicos y disímiles coberturas terrestres del entorno de la cuenca hidrográfica en imágenes de satélites.

2. Desarrollo

En investigaciones científicas relacionadas con las imágenes de satélites para lograr el procesamiento digital en las mismas se utiliza el análisis estadístico que son un sustento para analizar la calidad de la imagen, extraer las características espectrales particulares, valorar la correlación entre pares de canales, reducir el volumen de los datos, así como filtrar y segmentar la imagen, etc. tal como se muestra en la figura 1.

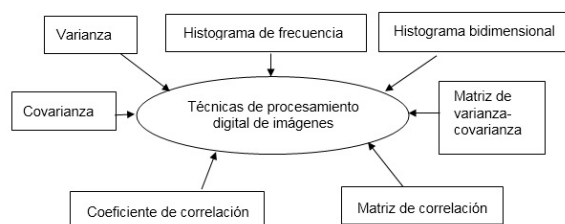


Figura 1. Técnicas de procesamiento digital de imágenes de satélites [10].

Autores como Lillesand y Kiefer, (1987); Harrison y Jupp [8]; López [10], Melh y Peinado [11], Basterra [2], Villar [15], corroboran la potencialidad de estas técnicas en el procesamiento digital de las imágenes de satélites y coinciden en que la representación visual de los datos de imagen necesita maximizar el contraste entre los niveles de visualización disponibles para proveer una diferenciación óptima de las características de la imagen. Existen dos métodos para llevar a cabo el re-escalado de una imagen estos se mencionan a continuación.

2.1 M3todo lineal

Los autores antes mencionados consideran que para mejorar el contraste total de la imagen, generalmente lo que se hace es seleccionar los valores m3nimo y m3ximo de la misma, auxili3ndose del histograma de frecuencias, posteriormente los valores m3nimo y m3ximo seleccionados son llevados a 0 y 255 respectivamente, realiz3ndose la operaci3n para el resto de los valores comprendidos entre los valores m3nimo y m3ximo.

Adem3s mencionan que se reportan buenos contrastes en las im3genes de tele-detecci3n seleccionando los valores m3nimo y m3ximo que representan el 1 % y el 99 % del histograma acumulativo de la imagen. Los valores que se encuentren por debajo del pixel que representa el 1 % y por encima del que represente el 99 % se pierden. Tambi3n se reporta el empleo de los pixeles que representen el 1 % y 95 %.

Para un intervalo de 0 a 255 el re-escalado quedar3a expresado como:

$$\begin{aligned} y &= 0, & x < x_{\min} \\ y &= 255 \frac{(x - x_{\min})}{(x_{\max} - x_{\min})}, & x_{\min} \leq x < x_{\max} \\ y &= 255, & x \geq x_{\max} \end{aligned}$$

donde y valor re-escalado; x valor a re-escalar.

2.2 M3todo no lineal

Seg3n L3pez [10] a diferencia del m3todo lineal en donde los valores son distribuidos proporcionalmente para llenar el intervalo de 0 a 255 aqu3 *selectivamente* se aumenta la separaci3n o distribuci3n entre los valores m3nimos o m3ximos de la imagen.

De manera general se puede implementar como una funci3n exponencial de la forma siguiente:

$$\begin{aligned} y &= A & x < x_{\min} \\ y &= A + B(x - x_{\min})^t & x_{\min} < x \leq x_{\max} \\ y &= A + B(x_{\max} - x_{\min}) & x > x_{\max} \end{aligned}$$

donde y valor re-escalado; x valor a re-escalar; A valor m3nimo de salida (puede ser 0 u otro); B factor de escala entre intervalo de entrada y de salida; t valor para el estimado.

Esta es una formulaci3n general y se pueden dar las siguientes situaciones: $t = 10$ Re-escalamiento Lineal; $t < 10$ Re-escalamiento no Lineal con funci3n logar3tmica, se aumenta el contraste entre los valores m3nimos de la imagen (zonas oscuras); $t > 10$ Re-escalamiento no Lineal con funci3n exponencial, se aumenta el contraste entre los valores m3ximos de la imagen (zonas claras).

De igual modo el contraste puede ser modificado aplicando a los pixeles de entrada otra funci3n matem3tica directamente programada: logaritmo, exponencial, etc.

Para la investigaci3n se asume el m3todo no lineal puesto que el histograma de las im3genes de sat3lites es multimodal, como es el caso de im3genes infrarrojas donde existe presencia de agua y vegetaci3n, se puede entonces re-escalar una caracter3stica en cuesti3n seleccionando sus valores m3nimo y m3ximo. En este caso el re-escalamiento de una caracter3stica

implica la negaci3n de la otra y viceversa. Luego que, con vistas a mejorar o resaltar la informaci3n que nos brindan las im3genes de sat3lites, es necesario realizar un proceso de filtraje en dichas im3genes

3. Filtraje de im3genes de sat3lites

De acuerdo con Villar [15] el proceso de filtraje de las im3genes (ver figura 2) se suele emplear principalmente para suavizar o eliminar ruidos en la imagen y para la detecci3n de bordes, se utilizan varios tipos de filtros.

Conociendo que una imagen queda representada por $f(x, y) = ND$ con una variaci3n espacial, se puede hacer corresponder dicha funci3n a una *se3al sinusoidal* con una frecuencia determinada (y , por tanto, con un per3odo T).

Si $f(x, y)$ es peri3dica, ser3 una funci3n infinita de senos y cosenos con diferentes amplitudes y fases. Las funciones $f(x, y) = f^\infty(\text{sen}, \text{cos})$, se conocen como *series de Fourier*.

Si $f(x, y)$ no es peri3dica, pero de variable real e integrable (pudiendo ser, a su vez, continua o discreta), se puede aplicar toda la teor3a de *transformadas de Fourier* y los procesos de *convoluci3n* y *laplacianos*.



Figura 2. Filtraje de im3genes de sat3lites [15].

Autores como L3pez [10] y Basterra [2] ratifican la utilizaci3n de estos filtros para mejorar la visualizaci3n en las im3genes de sat3lites. Adem3s Melh y Peinado [11] hacen referencia a otros filtros como: filtros de ensanches de bordes, filtros direccionales, filtros no lineales.

Luego de visualizar la informaci3n que contienen las im3genes de sat3lites es necesario extraer el conocimiento referente a la naturaleza espacial de las mismas mediante las t3cnicas de miner3a de datos.

3.1 M3todos utilizados en la miner3a de datos espaciales

En cuanto a los m3todos de miner3a de datos espaciales, en la investigaci3n que se propone realizar, se asumir3 el *m3todo de exploraci3n de asociaci3n espacial*, ya que el descubrimiento de reglas de asociaci3n espacial permite establecer, como su nombre lo indica, reglas que asocian objetos espaciales con uno o m3s objetos espaciales de su entorno. Desde el punto de vista conceptual, una regla de asociaci3n se define como $X \rightarrow Y$, donde X y Y son conjuntos de predicados espaciales o no espaciales (Vyas et al. [16]). En este sentido, se debe acotar que los predicados espaciales permiten calcular relaciones entre objetos y devuelven un valor booleano, entre los que se encuentran valores como *disjoint*, *touches*, *overlaps*, *contains*, *adjacent_to*, *near_by*, *inside*, *close_to*, entre otros.

En la técnica de asociación se introducen dos conceptos: mínimo soporte y mínima confianza. En las grandes bases de datos pueden encontrarse múltiples asociaciones entre los objetos, pero éstas deben poder aplicarse a pequeños grupos; por esto se deben filtrar las asociaciones utilizando mínimo soporte y mínima confianza [1].

Uno de los algoritmos más conocidos en la asociación espacial se denomina *A priori*, desarrollado por Agrawal et al [1], en 1993. Este algoritmo trabaja básicamente en dos pasos: en el primero, los grandes ítems son determinados de acuerdo con la frecuencia de los elementos dentro del grupo, mientras que en el segundo paso se detectan las reglas de asociación [1].

Como un complemento, las reglas de asociación espacial representan las relaciones objeto/predicado, las cuales contienen predicados espaciales.

3.2 Relaciones espaciales entre objetos geográficos. Criterios a tener en cuenta para establecer las reglas de asociaciones espaciales.

Según los autores Guillermo [6], Bogorny [3], Xiaoshenga [17], existen tres tipos de relaciones: distancia, dirección y topológicas (vea figura 3)

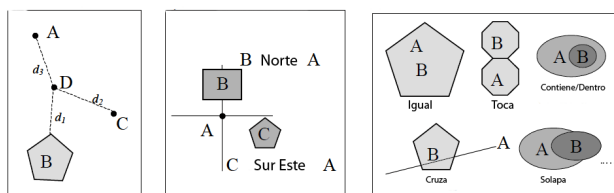


Figura 3. Ejemplo de relaciones espaciales [6].

En la tabla 1 se describe la clasificación de predicados espaciales según la categoría *Topológicos* que se asume en la investigación para establecer el grado de conectividad entre los distintos tipos de humedales y el entorno que lo circunda.

4. Taxonomía propuesta para los objetos hidrológicos presentes en las imágenes de satélites

Según Kubski [9] se define taxonomía como una jerarquía o superposición de relaciones entre las diferentes clases de un elemento. Estos elementos pueden clasificarse en distintas categorías. Así se establecen relaciones entre las mismas en términos de elementos hijos y elementos padres.

Para el desarrollo de la investigación propuesta, es necesario disponer a modo de datos de partida, de un esquema taxonómico general de los objetos hidrológicos que serán la base para el descubrimiento de las relaciones espaciales y no espaciales con el resto de los objetos ambientales que los circundan. Este esquema es muy importante, pues permitirá identificar los tipos de objetos hidrológicos presentes en el contexto de las imágenes de satélites desde una perspectiva homogénea para la investigación que se pretende realizar,

así como para investigaciones posteriores como punto de partida clasificar los mismos en las cuencas hidrográficas.

El esquema taxonómico de partida es general debido a que es necesario conocer los tipos de clasificaciones de objetos hidrológicos que están presentes a nivel nacional e internacional para de ahí generar nuevos esquemas taxonómicos de acuerdo con la imagen de satélite que se inserte en el SIG.

De manera preliminar y siguiendo los criterios de Montreux [12], Cowardin et al [5] y Cervantes [4], hemos adoptado la siguiente taxonomía que se muestra en el organigrama de la figura 4, con la cual se trabajará en la investigación que se llevará a cabo.

4.0.1 Conclusiones

Durante el proceso de investigación para este trabajo, se pudo corroborar la coincidencia de criterios de muchos investigadores al plantear que el proceso de generación y obtención de las reglas de asociación espacial es complejo y que requiere del empleo de diversas operaciones. Teniendo como referencia las limitaciones detectadas en los sistemas y métodos propuestos en diversas investigaciones realizadas por varios autores se arriba a las siguientes conclusiones:

1. Se considera apropiado a los efectos de la investigación que se llevará a cabo, utilizar el “método de exploración de asociaciones espaciales” como parte de la técnica de minería de datos, para el descubrimiento reglas de asociación espacial que relacionen a uno o más objetos espaciales. Esto se realizará utilizando tanto las imágenes de satélites como los SIG con el fin de utilizar estos datos para descubrir las asociaciones espaciales entre los objetos hidrológicos y las clases de cobertura que los circundan.
2. En cuanto a las relaciones espaciales entre objetos geográficos en la investigación, se propone utilizar las topológicas debido a que la representación de los datos geográficos es la vectorial ya que realiza una codificación eficiente de la topología y las operaciones espaciales.
3. En la investigación propuesta, se asumirá el esquema taxonómico de objetos hidrológicos en imágenes de satélites de observación de la tierra planteado anteriormente, el cual se utilizará ulteriormente como dato de partida para facilitar el descubrimiento de relaciones espaciales y no espaciales con los objetos ambientales del entorno en las imágenes de satélites del área de estudio, realizando la adecuación del esquema taxonómico general.

Referencias

- [1] Agrawal, R. T., Imielinski, Swami, A. Mining association rules between sets of items in large databases. Proceeding of the 1993 ACM SIGMOD International Conference on Management of Data, 1993. 1-4.

Tabla 1. Clasificación de predicados espaciales según la categoría Topológicos [13].

Categoría de Predicados	Operadores Espaciales	Ejemplo	Significado	Rango de Valores
Predicados Topológicos	separados	$\text{separados}(A, B)$	Objeto A separado del B	Verdadero, Falso
	intercepta	$\text{intercepta}(A, B)$	Objeto A intercepta al B	Verdadero, Falso
	contiene	$\text{contiene}(B, A)$	Objeto B contiene al A	Verdadero, Falso
	dentro	$\text{dentro}(A, B)$	Objeto A dentro del B	Verdadero, Falso
	topar	$\text{topar}(A, B)$	Objeto A toca el borde del B	Verdadero, Falso
	igual	$\text{igual}(A, B)$	Objeto A es igual al B	Verdadero, Falso
	cubre	$\text{cubre}(B, A)$	Objeto B cubre al A	Verdadero, Falso

- [2] Basterra, I. 1999. Procesamiento digital de imágenes.
- [3] Bogorny, V. 2006. Enhancing Spatial Association Rule Mining in Geographic Databases. Universidad Federal de Rio Grande del Sur.
- [4] Cervantes, M. 1994. Guía regional para el conocimiento, manejo y utilización de los humedales del noroeste de México., México.
- [5] Cowardin, L. M., Carter, V., Golet, F. C., Roe, E. T. L. 1979. Classification of wetlands and deepwater habitats of the United States. In: Program, B. S. (ed.) Fish and Wildlife Service. Washington, D.C.
- [6] Guillermo, O. 2010. Aplicación de minería de datos espaciales en un estudio de cambio de uso del suelo. Universidad de Concepción, Chile.
- [7] Han, J., Kamber, M. 2006. Data mining concepts and techniques, Morgan Kaufmann.
- [8] Harrison, B. A., Jupp., D. L. B. 1989. Introduction to Image Processing. Australia.
- [9] Kubski, M. I. 2005. Aplicación Orientada al Descubrimiento del Conocimiento en Bases de Datos. Universidad Nacional del Nordeste: Argentina.
- [10] López, J. 1990. Procesamiento digital de imágenes. Bogotá.
- [11] Melh, H., Peinado, O. 1992. In: Raumfahrt, D. Z. (ed.) Fundamentos del procesamiento digital de imágenes
- [12] Montreux 1990. Towards the Wise Use of Wetlands, Gran Bretaña.
- [13] Ronghua, M. 2005. Cognitive logic Representation of Spatial Association Rules of knowledge Discovery from GIS Database. 6.
- [14] Rosette, F. D. S. 2013. Método para la extracción de asociaciones espaciales en bases de datos geográficos. Doctorado, Ministerio de las Fuerzas Armadas Revolucionarias. Instituto Técnico Militar José Martí. Orden "Antonio Maceo". Orden "Carlos J. Finlay".
- [15] Villar, A. A. 2002. Teledetección y sistemas de tratamiento digital de imágenes.
- [16] Vyas, R., Kumar, L., Tiwary, U. 2007. Exploring spatial ARM (Spatial Association Rule Mining) for geo-decision support system. Journal of Computer Science, 3, 1-3
- [17] Xiaoshenga, L. 2005. Extract Spatial Association Rules by Method of Spatial Analysis. 4.

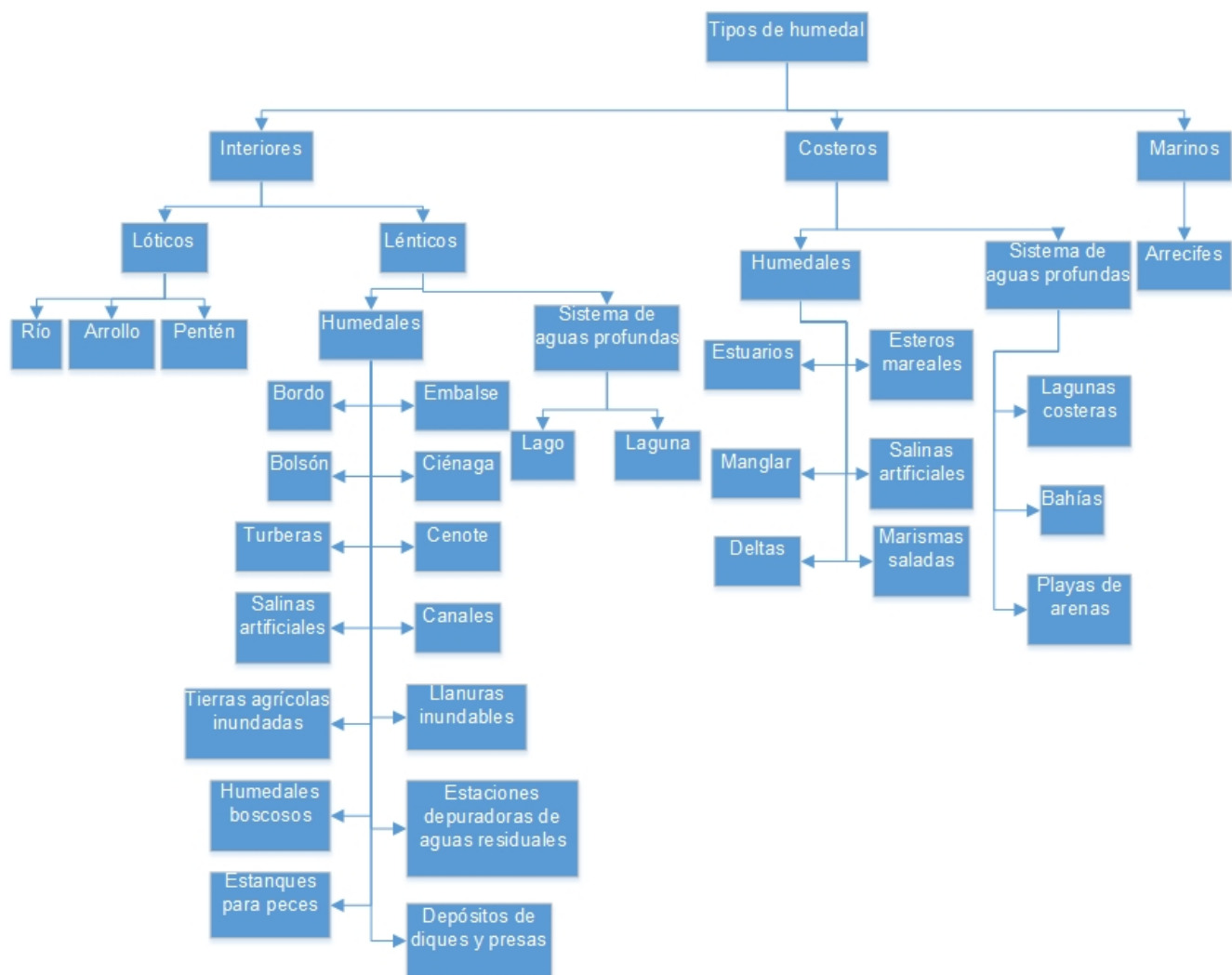


Figura 4. Taxonomía hidrológica a partir de las definiciones de MonTreux [12], Cowardin et al [5] y Cervantes [4].