

EL ANÁLISIS DE LOS DATOS EN EL SISTEMA DE INFORMACIÓN GEOGRÁFICA DE CUBA

Lic. Orlando Nouva Álvarez *

INTRUDUCCIÓN

El Instituto de Geografía de la Academia de Ciencias de Cuba ha desarrollado un Sistema de Información Geográfica (SIGC), cuyo objetivo central es la actualización sistemática de los principales temas tratados en el nuevo Atlas Nacional de Cuba (NANC).

Entre los objetivos particulares de este sistema está en posibilitar diferentes análisis de la información geográfica, así como conocer de forma rápida las características geográficas principales de un área, incluyendo sus condiciones topográficas.

Las bases de datos de los SIGC, que recogen diferentes formatos los variados datos que forman parte de la información geográfica dan la posibilidad de un acceso directo a su contenido por parte del usuario y en algunos casos, de un acceso selectivo mediante especificaciones limitada. Sin embargo los accesos selectivos complejos y los filtrados no están al alcance de las mismas.

El objetivo central de este trabajo es explicar la estructura del subsistema de análisis, el esquema de su funcionamiento, presentar el software y algunas muestras de sus aplicaciones: el subsistema de análisis del SIGC. Esquema funcional

La información sometida a análisis está almacenada en dos de las bases de datos con que cuenta el SIGC. Estas son: GEODATA, donde están registrados los atributos de los mapas del NANC captados espacialmente con el uso de una rejilla matricial con cuadrículas que abarcan aproximadamente 25 km² empleando el programa DIGICAPT para la captación mediante un tablero digitalizador, y GEOPUNTO, donde se encuentran los datos estadísticos referentes a fenómenos que se representan en los mapas mediante diagramas localizados puntualmente.

Desde el punto de vista funcional el subsistema de análisis del SIGC está compuesto por 4 módulos fundamentales que son: Análisis estadístico, Clasificación y Evaluación, Análisis Espacial y modelos Tridimensionales. En el primero se pueden realizar análisis estadísticos descriptivos así como inferencias con el sistema MICROSTAT y análisis factoriales con el programa COMPRIN. En el segundo módulo es posible hacer clasificaciones geográficas empleando varios métodos heurísticos de clasificación numérica, con el programa CLUSTER, evaluar un modelo de comportamiento de variables geográficas en un territorio y clasificar éste último de dependencia de su similitud al modelo, empleando el

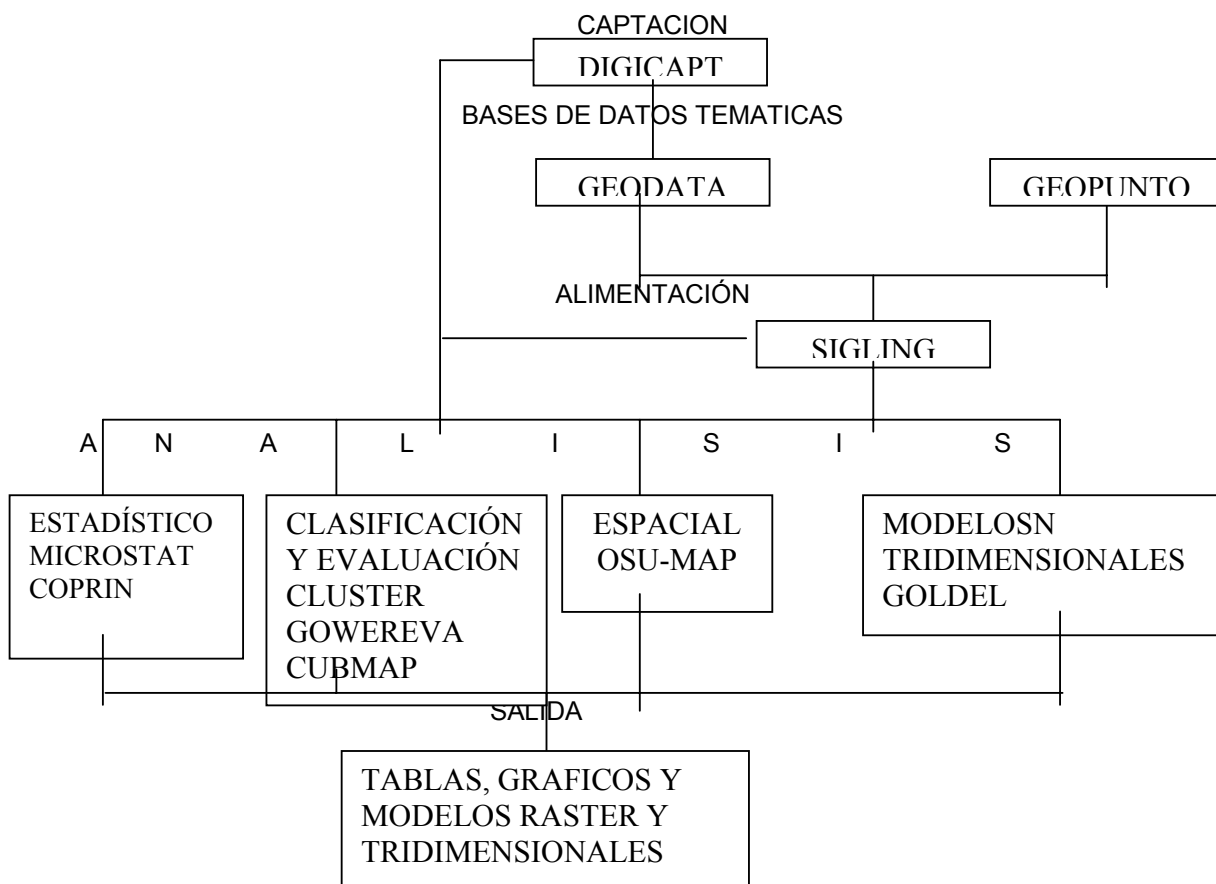
* Instituto de Geografía, Academia de Ciencias de Cuba.

coeficiente de similitud de Gower, con el programa GOWEREVA, o crear una tipología basada en discriminantes, clasificar un territorio en base a ella y después ir modificándola hasta lograr que sea optima, con el programa OSU-MAP, que basado en la cartografía Raster por sobreposición de matrices, permite análisis espaciales y la construcción de modelos cartográficos. Este sistema trabaja con la base geométrica Raster del SIGC en la que pueden almacenar datos de las bases planimétricas de cuba por provincias así como de mapas temáticos del NANC, contenidos en GEODATA.

En el cuarto módulo es posible construir modelos tridimensionales y de isolíneas, realizar cálculos volumétricos de sólidos y construir distintos tipos de gráficos entre otras opciones que brinda el sistema GOLDEN. Aquí se trabaja con una base de datos de puntos, a los cuales se les asocian, según un sistema de ejes cartesianos, valores en las coordenadas X, Y, así como el atributo correspondiente que se desea representar en la vertical o eje Z, la alineación a los módulos se procede mediante el programa SIGLINK

Los diferentes recorridos que presenta la información dentro del subsistema pueden apreciarse en la figura 1.

Figura 1 esquema funcional del subsistema de análisis del SIGC



Los sistemas y programas de análisis procesan información proveniente de las bases de datos GEODATA o GEOPUNTO según el caso. Los datos son seleccionados y extraídos de GEODATA por SIGLINK o exportados por GEOPUNTO y se preparan según mediante el alimentador, que se encuentra en un subdirectorío que lleva su nombre. Por su parte, el programa de capacitación de datos mediante el digitalizador, DIGICAPT, a la vez que capta la base de datos geométrica raster el sistema OSU-MAP y para la base de datos tridimensional que con que opera el sistema Golden. Estas bases son almacenadas en el subdirectorío SIGLIHK hasta el momento del análisis cuando el alimentador se encarga de trasladarlas el subdirectorío requerido. Por último, los resultados del análisis tienen por salida por la pantalla de la microcomputadora y pueden ser impresos, pudiendo ser en forma de tablas, gráficos y modelos cartográficos raster y tridimensionales, según sea el tipo de análisis y sus opciones de salida.

Tipos de análisis considerados. Programas y sistemas. Ejemplos

2.1 sistema MICROSTAT, marca registrada de Ecosoft INC versión 4.1, 1984.

Es un sistema de programas estadísticos que realiza las pruebas y procedimientos más comunes dentro de esta disciplina. Abarca tanto la estadística descriptiva como la inferencial.

Los métodos estadísticos se usan ampliamente en geografía para describir el comportamiento de muchos fenómenos de manera organizada y para formular y comprobar hipótesis teóricas. Entre los más empleados están los análisis de frecuencias, así como de los valores que permiten comparaciones entre distribuciones, los índices que miden el grado de relación entre las observaciones de varias variables y los análisis de regresión.

Programa COMPRIN, Facultad de Biología Universidad de la Habana, 1989

Permite reducir el número de variables que explican a un grupo de individuos objetos fenómenos, entidades especiales, con el mínimo de pérdida de la información, empleando para ello el análisis factorial.

Con este análisis se concentra un gran número de datos cuya manipulación es engorrosa, en un reducido número de variables produciendo una síntesis de información y evitando redundancias.

Programa CLUSTER, Instituto de Cibernética, Matemática, Física, Academia de Ciencias de Cuba, 1986.

Permite la clasificación jerárquica y agrupación de individuos (objetos, fenómenos, entidades espaciales), o de variables en niveles de semejanza, ejecutando varios métodos heurísticos para el reconocimiento de patrones no supervisados y teniendo en cuenta distancias medidas de semejanza.

Distancia euclídeana
Cuadrado de la distancia euclídeana
Distancia de Minkovski
Distancia media cuadrada
Distancia en tamaño
Distancia en forma
Coeficiente de Duran y Odell
Coeficiente de divergencia
Similitud por transformación
Coeficiente de Gower
Coseno del ángulo

Para la clasificación de variables están:

Coeficiente de correlación
Coseno del ángulo
Covarianza
Distancia euclídeana

Los métodos de clasificación son:

Conexión simple
Conexión compleja
Centroide
Mediana
Premediación entre los grupos formados
Premediación dentro del nuevo grupo
Ward

Un ejemplo de aplicación del programa CLUSTER es la clasificación de 32 cuencas fluviales en grupos semejantes, en base a 9 variables hidrométricas escogidas. La medida de semejanza utilizada es la distancia euclídeana y del método de clasificación es el de WARD.

2.4 Programa GOWEREVA, Instituto de Geografía, Academia de Ciencias de Cuba, versión 3.0 1990

Permite clasificar individuos (objetos, fenómenos, entidades espaciales), en intervalos de similitud con respecto a un individuo modelo conocido, empleando la distancia de Gower, las variables analizadas que dan evaluadas según el modelo en cada uno de los entes restantes.

Esta metodología de aplicación de la distancia de Gower ha sido experimentada en la regionalización de territorios según sus condiciones naturales para el desarrollo de actividades económicas, aunque tiene una aplicación más general.

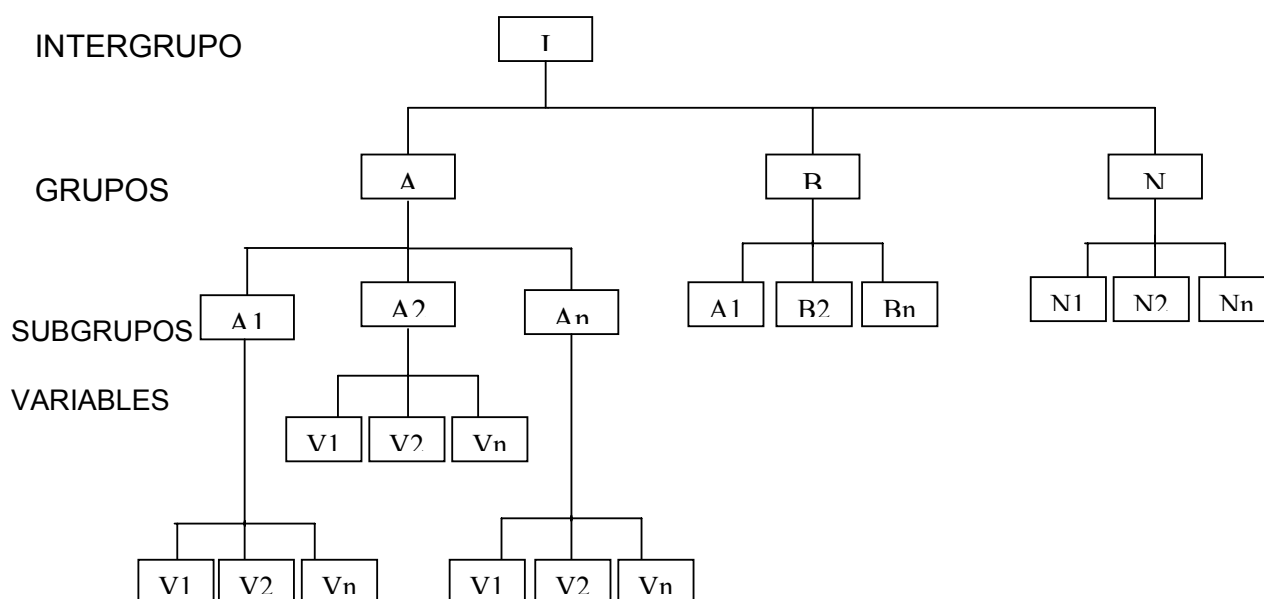
Para ello, determina un modelo óptimo de condiciones naturales para el desarrollo de la actividad que se trata, teniendo en cuenta las exigencias que el desarrollo de esa actividad necesita del medio natural. Después se determina la distancia Gower existente entre el modelo y cada uno de los individuos espaciales del territorio de análisis, caracterizados por las mismas variables del modelo pero con datos reales.

Se analiza a continuación la frecuencia de los valores de la distancia Gower para conocer como se agrupan estos y determinan intervalos de agrupación con lo cual los individuos quedan clasificados. Por último, creando una leyenda tipológica se obtiene como resultado un mapa del territorio regionalizado.

Programa CUBMAP, Instituto de Geografía, Academia de Ciencias de Cuba, versión 4.0, 1990.

Permite hacer regionalizaciones a partir de una tipología o patrón de clasificación concebida por el usuario, compuesta por subgrupos e intergrupos, dados por los componentes de varias variables. Emplea el análisis de discriminantes (figura 2).

Figura 2. Estructura de una tipología



Además de la tipología puede ir modificándose en dependencia de los resultados de la clasificación de los resultados de la clasificación del territorio en base a ello, por lo que esta queda a su vez optimizada.

System OSU-MAP Geographic Information systems Laboratory Department of Geography, The Ohio State University, version 1.0 1987.

Es un sistema de programas que, basado en la cartografía raster por sobreposición de matrices, permite el análisis espacial y la construcción de modelos cartográficos de búsqueda, síntesis y relaciones espaciales en general.

Los mapas llevados a matrices o también llamados sobrepuestas pueden someterse a múltiples operaciones que pueden dividirse en 4 categorías principales, las reclasificaciones de los atributos de las superposiciones superpuestas la determinación de distancias y la de conectividad entre individuos espaciales, dada por variados factores y la caracterización de las vecindades cartográficas.

Sistema GOLDEN, Golden Software Inc. Versión 3.0 1987.

Es un sistema de programas gráficos que permite la construcción de modelos tridimensionales y de isolíneas, así como la elaboración de gráficos de diferentes tipos e incluye una amplia gama de opciones para el tratamiento de la información antes de su representación final. Se realizan con él, cálculos de volúmenes sólidos y también pueden trazarse perfiles.

Programa Alimentador SIGLINK Instituto de Geografía Academia de Ciencias de Cuba, 1989.

Permite extraer de las bases de datos, la información seleccionada por el usuario y prepararla en el formato necesario para su entrada en los distintos programas y sistemas de análisis, es el conector entre las bases y los programas de aplicación.

Es importante señalar que todos estos programas y sistemas anteriores citados tanto como parte del SIGC, así como en módulos independientes. De esta forma los individuos espaciales declarados para el análisis pueden ser, los establecidos por el SIGC cuando se trabaja con sus datos o los convenidos por cualquier usuario en particular cuando se trabaja de forma independiente.

BIBLIOGRAFÍA

Díaz L. R., Candeaux R., y Nouva O., 1990 Sistema de Información Geográfica de Cuba, Academia de Ciencias de Cuba.

Figura no.3 Ejemplo de salida de un análisis estadístico con el programa MICROSTAT.

```

----- REGRESSION ANALYSIS -----
HEADER DATA FOR: C:VARH LABEL: Correlacion multiple
NUMBER OF CASES: 58 NUMBER OF VARIABLES: 3
-----
INDEX NAME MEAN STD. DEV.
1 Ma 14.9291 9.7437
2 A 402.8528 745.8169
DEP. VAR.: DSTa 1.8576 2.1810
-----
DEPENDENT VARIABLE: DSTa

VAR. REGRESSION COEFFICIENT ADJUSTED R SQUARED = .7617
Ma .0140 R SQUARED = .7701
A .0026 MULTIPLE R = .8775
CONSTANT .5887
-----

```

Figura no.4 Ejemplo de salida de una clasificación con el programa CLUSTER.

```

CLUSTER.

Jobabo - Jobabo, 21--|-----|
Chaparra-ElRoble 22--|-----|
Sevilla-LaFortal 20--|-----|
Buant.-Sta.Sofia 26--|-----|
Bihara - Jobo 23--|-----|
Cayojo-cayojo 27--|-----|
Caazan-Limoncit 24--|-----|
Damuji - Rodas, 16--|-----|
Zaza-PasoVentura 18--|-----|
Agabama - Sopimpa 17--|-----|
Mayari-La Emajag 25--|-----|
Salado-San.Carlo 28--|-----|
Taco-Taco-ElJard 30--|-----|
Sta.Cruz-Sta.Ana 31--|-----|
Toa - El toro, 13--|-----|
Buey -San Miguel 32--|-----|
Baconao -Trucucu 9--|-----|
Veteras-Palenqui 12--|-----|
Higuani-PteCarrt 8--|-----|
SaguadeI.-ElInfi 10--|-----|
Bano - Sta. Rosa 11--|-----|
Canas-SanAgustin 15--|-----|
Jatib.Nort-ElRio 19--|-----|
Canas - Canas 14--|-----|
Ceimito-El Centr 6--|-----|
Bacunagua-StoDow 29--|-----|
Sani.yeartz-Elit 4--|-----|
Cuyagt-V anivers 1--|-----|
Hondo - Pilotes 5--|-----|
San Diego-LosGav 7--|-----|
Cuyagt- La Guira 2--|-----|
Cuyagt-Portales 3--|-----|

Individuos No. 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

```

Figura no.5 Ejemplo de evaluación y regionalización con el programa GOWEREVA.

Evaluación para la agricultura

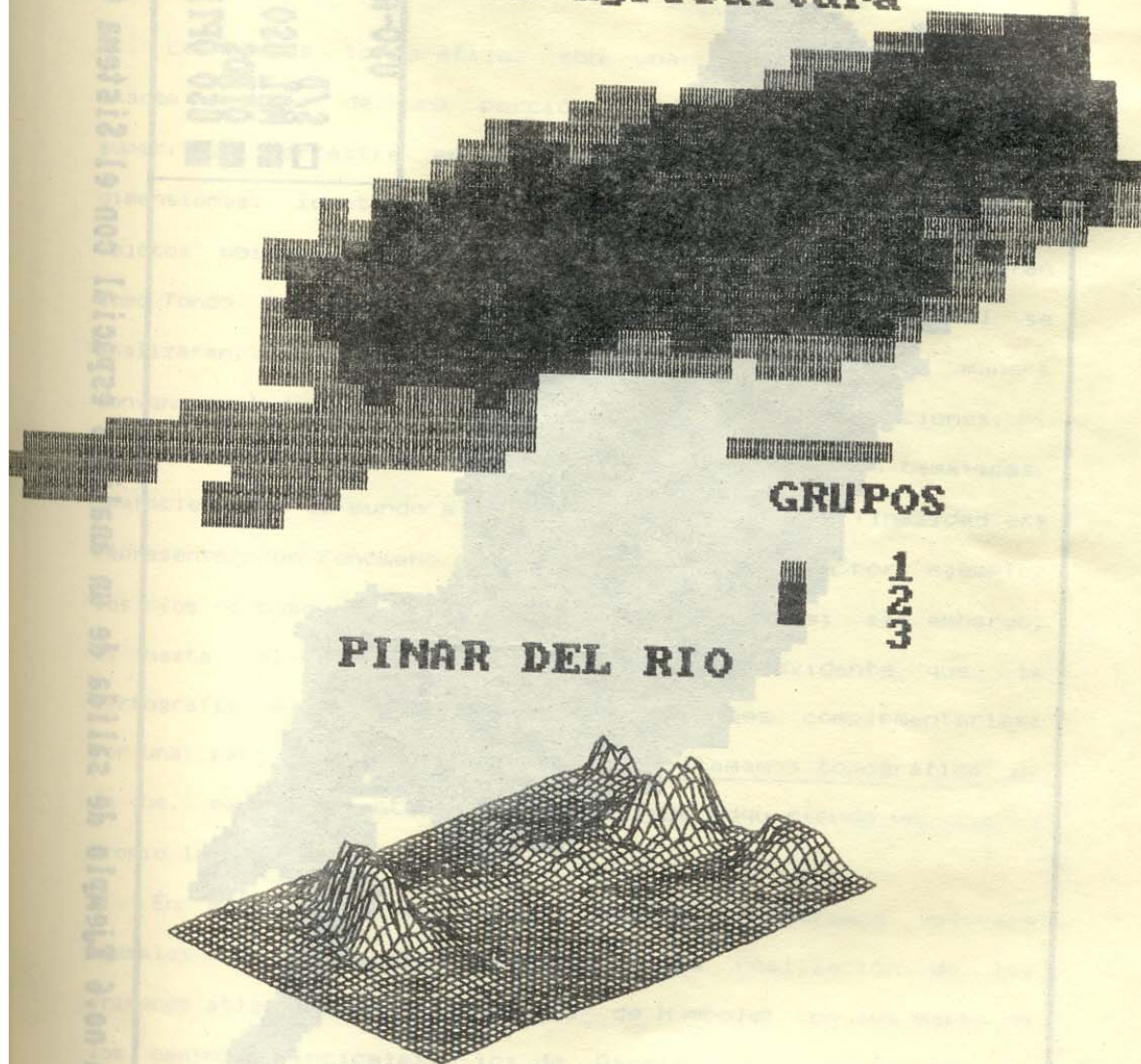


Figura no.7 Modelo tridimensional

