



Instituto Interamericano de
Cooperación para la Agricultura



UNIVERSIDAD DE
COSTA RICA



Instituto Nacional de Estadística
y Censos de Costa Rica



Observatorio del
Desarrollo



Centro Centroamericano
de Población



Escuela de Geografía

LO RURAL ES DIVERSO: EVIDENCIA PARA EL CASO DE COSTA RICA

Adrián Rodríguez y Milagro Saborío
Editores

*San José, Costa Rica,
Junio, 2008*

© Instituto Interamericano de Cooperación para la Agricultura (IICA). 2008

El Instituto promueve el uso justo de este documento. Se solicita que sea citado apropiadamente cuando corresponda

Esta publicación también está disponible en formato electrónico (PDF) en el sitio Web institucional en www.iica.int.

Coordinación editorial: Adrián Rodríguez y Milagro Saborío
Diagramado y diseño de portada: Ana Catalina Lizano
Impresión: Imprenta IICA

Lo rural es diverso: evidencia para el caso de Costa Rica / ed. por
Adrián Rodríguez, Milagro Saborío – San José, C.R. : IICA, 2008.
208 p.; 19cm x 27cm.

ISBN13: 978-92-9039-855-4

1. Desarrollo rural – Costa Rica 2. Censos de población 3. Zonas urbanas 4. Zonas rurales 5. Costa Rica I. Rodríguez, Adrián II. Saborío, Milagro III. IICA IV. Título

AGRIS
E50

DEWEY
307.72097286

San José, Costa Rica
2008

RASTERIZACIÓN DE LOS CENSOS DE POBLACIÓN Y VIVIENDA DE 1973, 1984 Y 2000

Róger Bonilla y Luis Rosero

1. INTRODUCCIÓN

Este documento presenta la metodología y resultados del proceso de “rasterización” espacial de los Censos de Población y Vivienda de Costa Rica, para los años 1973, 1984 y 2000. Se entiende por rasterización a la estimación de la distribución espacial de la población y las viviendas de esos censos, en una cuadrícula, enrejado o “grid” de celdas de igual tamaño, sobre todo el territorio nacional. En el estudio se usó un tamaño de celda de 500 metros de lado. Eso significa una división del país en aproximadamente 200.000 de esas celdas. La representación espacial de la población en esta clase de celdas o cuadrículas es equivalente a crear una “superficie” de población o de viviendas. El objetivo principal del ejercicio de rasterización es hacer posible la comparación y el estudio de la información censal (personas y viviendas) con un alto grado de desagregación geográfica.

La información censal en Costa Rica se recolecta a partir de una unidad denominada *segmento censal*. Posteriormente dicha información se puede organizar por las divisiones político-administrativas, como distritos, cantones y provincias, u otros criterios de agregación, tales como regiones de planificación. Para representar la distribución espacial de la población generalmente se usan mapas, compuestos por unidades geográficas de formas y tamaños irregulares,

correspondientes a las divisiones político-administrativas indicadas. A esas unidades administrativas se les denomina polígonos. El uso de polígonos irregulares, definidos por conveniencia administrativa, presenta muchos problemas para el análisis geográfico, en particular para representar patrones relacionados a la densidad poblacional (Martín, 1999). También presenta problemas para hacer comparaciones en el tiempo, debido al frecuente cambio de los límites de esas unidades, especialmente de las más desagregadas¹.

La alternativa a la representación espacial de la población en polígonos irregulares es la representación "ráster", en celdas de igual forma y tamaño. Esta representación crea una superficie poblacional. La idea de que la población pueda ser representada más adecuadamente y modelada como una superficie no es nueva. Autores como Schmid y MacCannell (1955), Nordbeck y Rystedt (1970), Tobler (1979) y Goodchild *et al.* (1993) han interpolado la población en una superficie utilizando diferentes enfoques.

Este informe presenta el procedimiento seguido para estimar la cantidad de personas y de viviendas existentes en cada una de las parcelas cuadradas de 500 metros de lado en que se divide al país. La estimación se efectúa para los años censales de 1973, 1984 y 2000, a partir de información *geocodificada* (localizada espacialmente con coordenadas geográficas) de los segmentos censales respectivos.

Los objetivos específicos de este estudio son: a) desarrollar y validar una metodología para la rasterización de los censos de población y vivienda; b) producir una base de datos con aproximadamente 200 mil celdas, de 500 metros de lado, con información censal seleccionada (población y vivienda); c) representar esta información en mapas; d) determinar parámetros con que se ha definido en los censos la población urbana - rural; y e) examinar escenarios alternativos de definiciones urbano - rural. A continuación, el desarrollo de la metodología se presenta en la Sección 2 y los resultados y conclusiones en la Sección 3.

1 Por ejemplo, desde el último Censo de Población, en el año 2000, se han creado 9 nuevos distritos.

2. DATOS Y MÉTODOS

2.1 DATOS Y FUENTES DE INFORMACIÓN

La información demográfica de los censos de población de 1973, 1984 y 2000 se integró en un sistema de información geográfico (SIG) por medio de cerca de 18.000 puntos en el territorio, que representan la localización de los centroides de los segmentos censales para el año 2000, 11.000 puntos para 1984 y 5.000 puntos para 1973. Para ello fue necesario geocodificar (establecer las coordenadas geográficas de latitud y longitud) de cada uno de estos puntos. La geocodificación se efectuó siguiendo el procedimiento propuesto por Rosero y Palloni (1998), sobre cartografía usada en el censo y que está disponible en el Instituto Nacional de Estadística y Censos (INEC).

Un segmento censal es una pequeña área geográfica bien delimitada que incluye aproximadamente 60 viviendas en el área urbana y 40 viviendas en el área rural. El INEC define los segmentos censales con el propósito de asignar las áreas de enumeración a los empadronadores. El segmento suele ser de una o dos manzanas en áreas urbanas. En las áreas rurales, en cambio, puede alcanzar varios centenares de hectáreas de superficie, aunque las viviendas tienden a concentrarse en una fracción de este territorio.

Los segmentos se geocodificaron como puntos y no como polígonos. Para ello el primer paso fue establecer un centroide poblacional del segmento. Este centroide es un punto imaginario equidistante a cada una de las viviendas del segmento. Luego se identificaron en el mapa las coordenadas cartesianas (latitud y longitud) del centroide, para lo cual fue necesario apoyarse en puntos de referencia adicionales, debido a que la cartografía censal no está georeferenciada (no tiene las coordenadas terrestres). El procedimiento supone inicialmente que todas las viviendas del segmento censal se concentran en un solo punto. Luego, con la rasterización, se remueve este supuesto y se distribuye a la población y a las viviendas en toda la superficie.

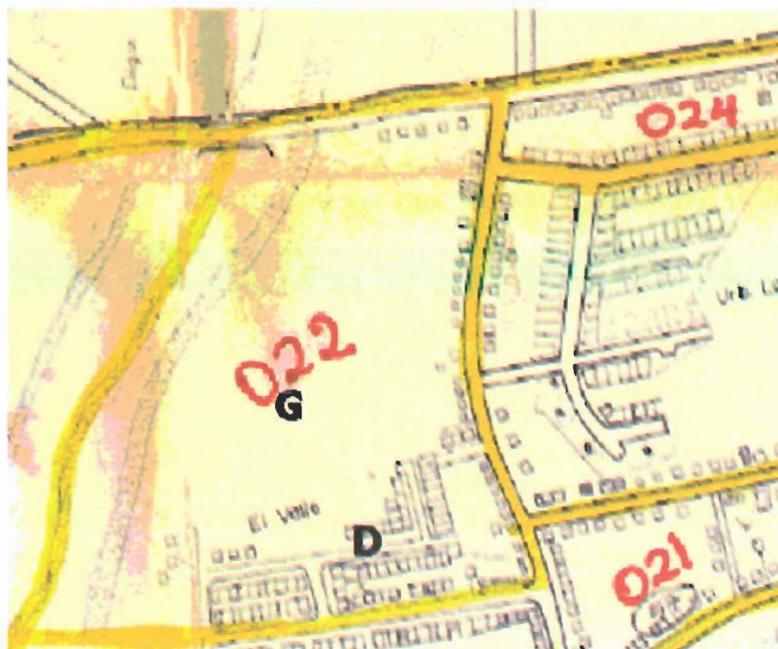
La Figura 1 muestra un segmento urbano real del censo del 2000. Lo ideal sería representar la información de este mapa con coordenadas para cada una de las viviendas. Ello, aparte de la dificultad y costo de llevarlo a cabo², tiene el inconveniente de que podría violar el principio de confidencialidad de los datos

2 En Costa Rica eso se agrava por la inexistencia de una nomenclatura de calles y números de casa.

estadísticos. La información debe, por tanto, agregarse y la unidad lógica más pequeña de agregación es el segmento censal. Para representar espacialmente la información de los segmentos, en primera instancia parecería lógico hacerlo con polígonos. Sin embargo, al proceder así, implícitamente se asume que las viviendas y la población se distribuyen homogéneamente dentro del polígono. La gráfica muestra que tal cosa no ocurre en la realidad: las viviendas tienden a concentrarse en una fracción del polígono. Y en las zonas rurales la ausencia de una distribución uniforme es aún mayor. Ante ello es preferible para ciertos usos representar espacialmente al segmento por un centroide demográfico, como el indicado en la gráfica. El centroide demográfico representará mejor la realidad que el centroide geográfico del polígono, como se muestra en la figura. Bracken (1989) muestra para el censo de Inglaterra las ventajas de los centroides demográficos para la representación espacial de la población.

Figura 1

Un segmento censal y su centroide demográfico (D) y Geográfico (G).



Las coordenadas geográficas de los centroides demográficos se agregaron a una base de datos de segmentos con información de los censos de población. Esta base de datos fue creada a partir del archivo electrónico original de los censos de población proporcionado por el INEC y que puede ser consultado por Internet en el Centro Centroamericano de Población (<http://censos.ccp.ucr.ac.cr>).

2.2 METODOLOGÍA DE RASTERIZACIÓN

El proceso de rasterización de la población consiste en “interpolar” la población, asignando ponderadores de distancia a celdas uniformes del ráster del SIG. El objetivo es redistribuir la población de los centroides a las áreas de donde fueron originalmente tomados. El resultado del proceso produce una matriz de celdas con valores estimados de población, cuya suma es el total de la población representado en los valores de la población en los centroides de los segmentos censales.

El modelo reconstruye la geografía de la población de una forma que sea independiente de la ubicación de los límites arbitrarios de los segmentos censales. La Figura 2 ilustra, a la izquierda, los segmentos censales y los centroides demográficos, los cuales contienen información poblacional. A la derecha de la gráfica se ilustra el resultado del proceso de rasterización o “expandimiento” de la información contenida en un punto hacia las celdas aledañas. En las zonas negras la información poblacional tiene valores de cero. El gradiente de los tonos de verde muestra los valores de la rasterización. Los tonos más claros (blancos y verdes claros) son valores bajos de la población rasterizada. Entre más se acerca al verde fuerte es porque el método de rasterización asignó más población en dichas celdas.

Figura 2

Ilustración de los segmentos censales y centroides poblacionales y el resultado de la rasterización.

Segmentos censales y centroides



Resultado de la rasterización



La población en los centroides de los segmentos censales se distribuye en las celdas aledañas y muchas celdas recibirán población de más de un centroide. De forma inversa, la mayoría de las celdas del ráster que están fuera del alcance de los centroides estarán sin población (Bracken y Martín, 1989).

La técnica de redistribución de la población depende de dos parámetros: a) el tamaño del radio de interpolación (r), el cual determina el área en donde la población se expandirá; y b) el parámetro α de la función del decaimiento de la distancia, la cual define el "suavizamiento" de la superficie dentro de las áreas pobladas.

El algoritmo de redistribución pasa por cada centroide, cuya cobertura está determinada por la densidad local de centroides. Los ponderadores se asignan a partir de las celdas contiguas, las cuales reciben una proporción del conteo del centroide. La población de cada celda P_i está estimada por:

$$\hat{P}_i = \sum_{j=1}^N P_j W_{ij} \quad (1)$$

en donde P_j es la población del centroide j y W_{ij} es el ponderador del centroide j con respecto a la celda i del ráster. Sólo los centroides que caen dentro del radio de interpolación son evaluados y el tamaño de dicho radio determina cuántos centroides serán usados para calcular la distancia promedio entre los centroides (d). En los modelos usados por Bracken y Martín (1989), los ponderadores están determinados por:

$$W_{ij} = \left[\frac{d^2 - s^2}{d^2 + s^2} \right]^\alpha \quad (2)$$

en donde s es la distancia entre la celda i del ráster y el centroide del segmento censal j .

La rasterización fue realizada con el programa SurfaceBuilder versión 0.1.0 (Martín, 1999). Los mapas fueron representados con el programa MapInfo versión 5 (MapInfo Corporation, 1985). La base de datos resultante fue generada en formato DBF III y luego analizada con el paquete estadístico Stata.

2.3 VALIDACIÓN

Para validar el procedimiento de rasterización se compararon los resultados obtenidos con conteos reales de viviendas en una muestra de celdas urbanas y otra de celdas rurales. Una alta coincidencia entre los resultados de la rasterización y los datos reales de campo, medida por el coeficiente de correlación de Pearson (Johnson, 2000) entre los conteos de viviendas por celda de las dos series, indicaría que el modelo utilizado está reconstruyendo la geografía de la población de una forma apropiada. Para efectuar esta comparación, los datos "reales" se obtuvieron por georeferenciación de cada una de las viviendas en la muestra. Luego se superpuso a esta capa de información geográfica una cuadrícula con los límites de cada celda de 500 m y se procedió al conteo de las viviendas dentro de cada celda.

La validación en la zona urbana se realizó en el distrito de San Pablo de Heredia y zonas aledañas. La zona rural fue constituida por los distritos de Sierpe y Puerto Jiménez, que conforman la península de Osa. Para la comparación, en la zona urbana se contó con 24 celdas y 1.825 ubicaciones geográficas de viviendas. En la zona rural hubo 1309 celdas y 2.060 viviendas. Las ubicaciones geográficas de estas cerca de 4.000 viviendas se habían efectuado en el CCP para otros estudios con la cartografía del censo de 1984. La validación se hizo, en consecuencia, contra la rasterización de dicho censo en las áreas de las dos muestras. El ejercicio de validación sirvió, además, para identificar los valores óptimos de los parámetros correspondientes al radio y al efecto de la distancia, para usar en la rasterización de los censos.

3. RESULTADOS

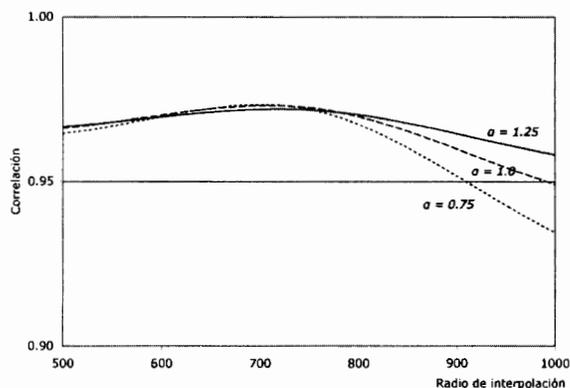
3.1 VALIDACIÓN

La Figura 3 muestra la correlación entre el conteo real de las viviendas en la muestra de celdas y la estimación por rasterización a partir de los centroides poblacionales de los segmentos censales. Se observa, en primer lugar, que el coeficiente de correlación entre las dos series llega a ser de 0,97 en el área urbana y 0,84 en el área rural; es decir, que el procedimiento de rasterización reproduce bastante bien conteos reales de viviendas en celdas de 500 metros (este resultado puede variar para celdas de diferente tamaño con correlaciones más altas para celdas más grandes y correlaciones más bajas para celdas más desagregadas).

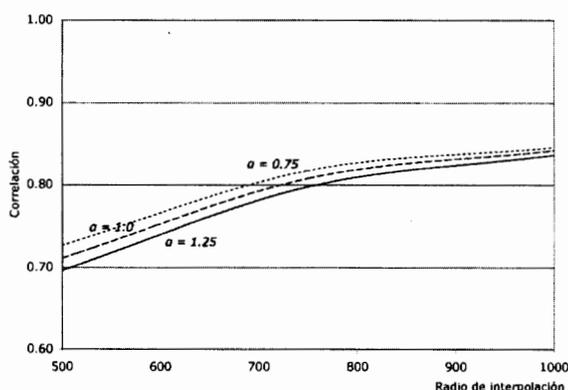
Figura 3

Correlación entre el conteo real de la población y el resultado de la población rasterizada. Costa Rica (distritos experimentales).

San Pablo de Heredia (Urbano)



Sierpe y Puerto Jiménez (Rural)



En segundo lugar, la Figura 3 muestra que se alcanza un óptimo de correlación máxima fijando el parámetro de distancia en $\alpha = 1$ y el radio de interpolación en 750 metros. Los valores anteriores corresponden al área urbana. En el área rural, el óptimo está dado por un $\alpha = 0.75$, pero la ganancia sobre el $\alpha = 1$ no es muy importante, por lo que se opta por usar un único parámetro de uno para todas las celdas. Para el radio se observa en el área rural que a partir de los 750 metros las ganancias son mínimas, por lo que también se opta por el valor encontrado en el área urbana. Los parámetros seleccionados para rasterizar la población de los segmentos censales de Costa Rica coinciden con los usados por Bracken y Martín en Inglaterra, el país de Gales y Escocia (1995).

3.2 SUPERFICIES ESTIMADAS

Las figuras 4, 5 y 6 resumen en mapas el resultado de la rasterización de las viviendas de Costa Rica para los censos 1973, 1984 y 2000 respectivamente. Las figuras 7, 8 y 9 presentan la resultados para la población. El Cuadro 1 ilustra una sección de la base de datos. En ella, las celdas están identificadas por las coordenadas del centro de esta, proyección Costa Rica Lambert Norte.

Las superficies de vivienda y población estimadas se pueden leer desde dos puntos de vista: de acuerdo con el territorio o según como se distribuye la

Cuadro 1

Ilustración de una sección de las variables de la base de datos.

Longitud (x)	Latitud (y)	Viviendas 1973	Viviendas 1984	Viviendas 2000	Población 1973	Población 1984	Población 2000
513250	220250	19	38	75	109	186	286
513250	220750	90	92	128	506	413	456
513250	221250	120	102	207	618	417	721
513250	221750	366	418	548	1712	1666	1725
513250	222250	727	720	729	3349	2731	2053
513250	222750	411	400	406	1914	1577	1235
513250	223250	101	52	79	526	226	255
513250	223750	33	6	33	194	24	106
513250	224250	158	144	63	972	565	243

población. El Cuadro 2 resume las dos opciones. Desde el punto de vista del territorio, la gran mayoría del país está deshabitada: el 83% tiene cero población según el censo del 2000; en 1973 ese porcentaje era del 92%. Si a lo anterior se agrega el territorio con densidad baja (de menos de 50 habitantes por celda o 200 por km²) se llega al 95% de la superficie del país. En el otro extremo, el territorio de Costa Rica con densidades de 1.000 o más habitantes por celda (4.000 habitantes por km²) no llega al 1%, pero allí reside el 41% de la población según el censo del 2000 (33% en 1973 y 36% en 1984). La distribución de la población en el territorio es, por tanto, extremadamente desigual.

Cuadro 2

Costa Rica: distribución del territorio y la población, por censo, según densidad de la celda

Población por celda de 500m	Territorio porcentaje			Población porcentaje		
	1973	1984	2000	1973	1984	2000
Total	100	100	100	100	100	100
0	92.4	86.6	82.5	0	0	0
1 a 49	3.8	9.4	12.1	9	16	13
50 a 99	1.9	2.3	2.7	15	13	10
100 a 249	1.3	0.9	1.4	20	11	11
250 a 999	0.4	0.6	1.0	23	24	26
1000 a 1999	0.1	0.2	0.3	15	21	22
2000 o más	0.1	0.1	0.1	18	15	19

Un objetivo de este análisis espacial de la población es aportar elementos para una mejor definición de lo urbano y lo rural. La manera más rudimentaria o un primer paso para diferenciar lo rural de lo urbano es con base en la densidad de población. Las superficies censales generadas en este estudio permiten dar este primer paso muy fácilmente. El Cuadro 3 muestra porcentajes hipotéticos de población urbana en los distintos censos si esta fuese definida con base a la densidad demográfica exclusivamente. Por ejemplo, si se tomase como punto de corte los 1.000 habitantes por km², el censo del 2000 clasificaría al 66% de la población como urbana. Si este valor se sube a 2.000 habitantes por km² el porcentaje de población urbana sería 55%. De acuerdo con la definición de urbano aplicada en el Censo de Población, esta proporción fue de 59% en el 2.000. El punto de corte que aproximadamente reproduce este porcentaje es, por tanto algo menos de los 2.000 habitantes por km².

Cuadro 3

Costa Rica, porcentaje de población urbana, por censo, según definición por densidad.

Densidad por km ²	1973	1984	2000
1000 hab. o más	55	60	66
2000 hab. o más	46	49	55
4000 hab. o más	33	36	41
8000 hab. o más	18	15	19
En censos	41	44	59

Si se toma como urbana la población con densidades de 2.000 o más hab/ km², ésta habría sido 46% en 1973 y 49% en 1984, valores por encima del porcentaje urbano de acuerdo a la definición censal, que fue 41% y 44%. Pareciera que la definición de lo urbano en esos censos dejó por fuera zonas importantes de alta densidad.

Los comentarios anteriores son, desde luego, análisis muy preliminares sobre la clasificación urbana. Al criterio de densidad, único utilizado aquí, deben agregarse otros criterios como infraestructura física, disponibilidad de ciertos servicios y, sobre todo, el de masa crítica mínima para que un lugar poblado sea considerado una urbe.

Figura 4

Costa Rica: mapa del número de viviendas rasterizado, Censo 1973.

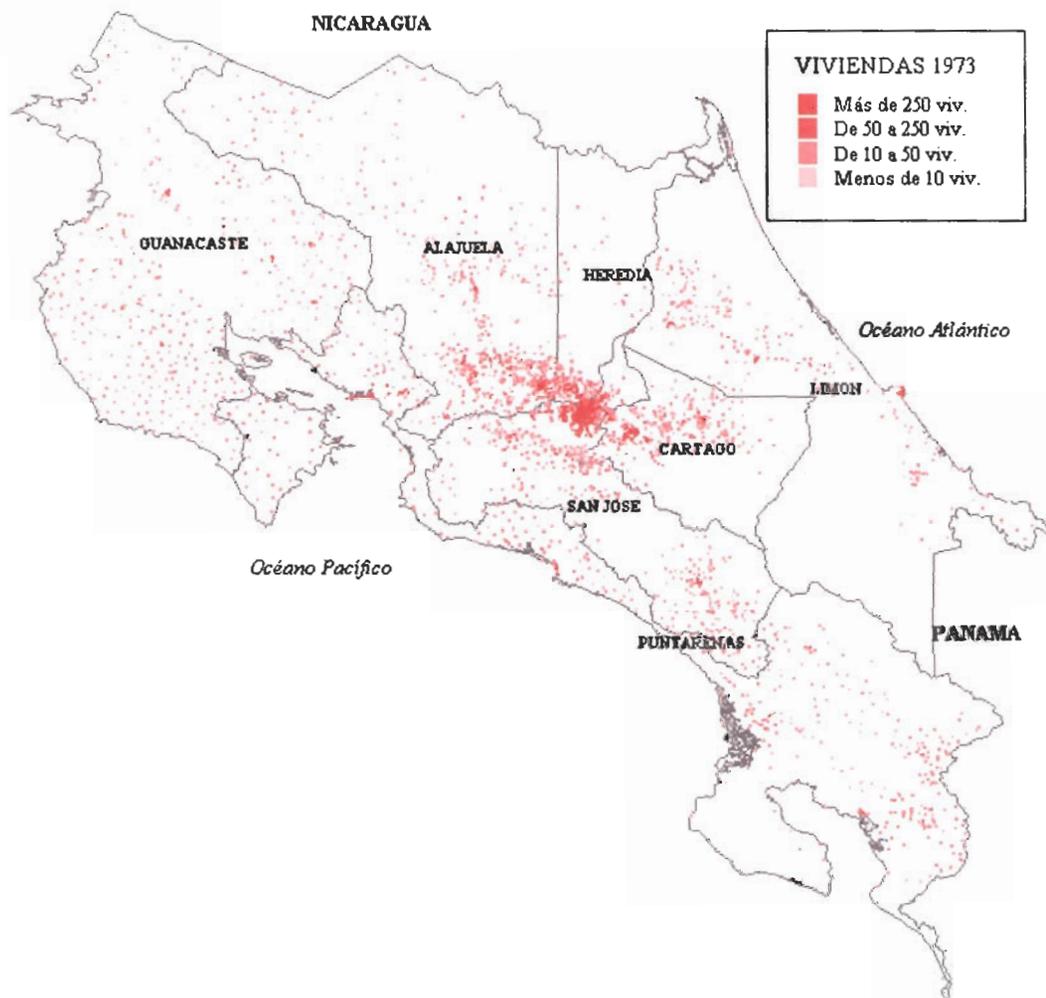


Figura 5

Costa Rica: mapa del número de viviendas rasterizado, Censo 1984.

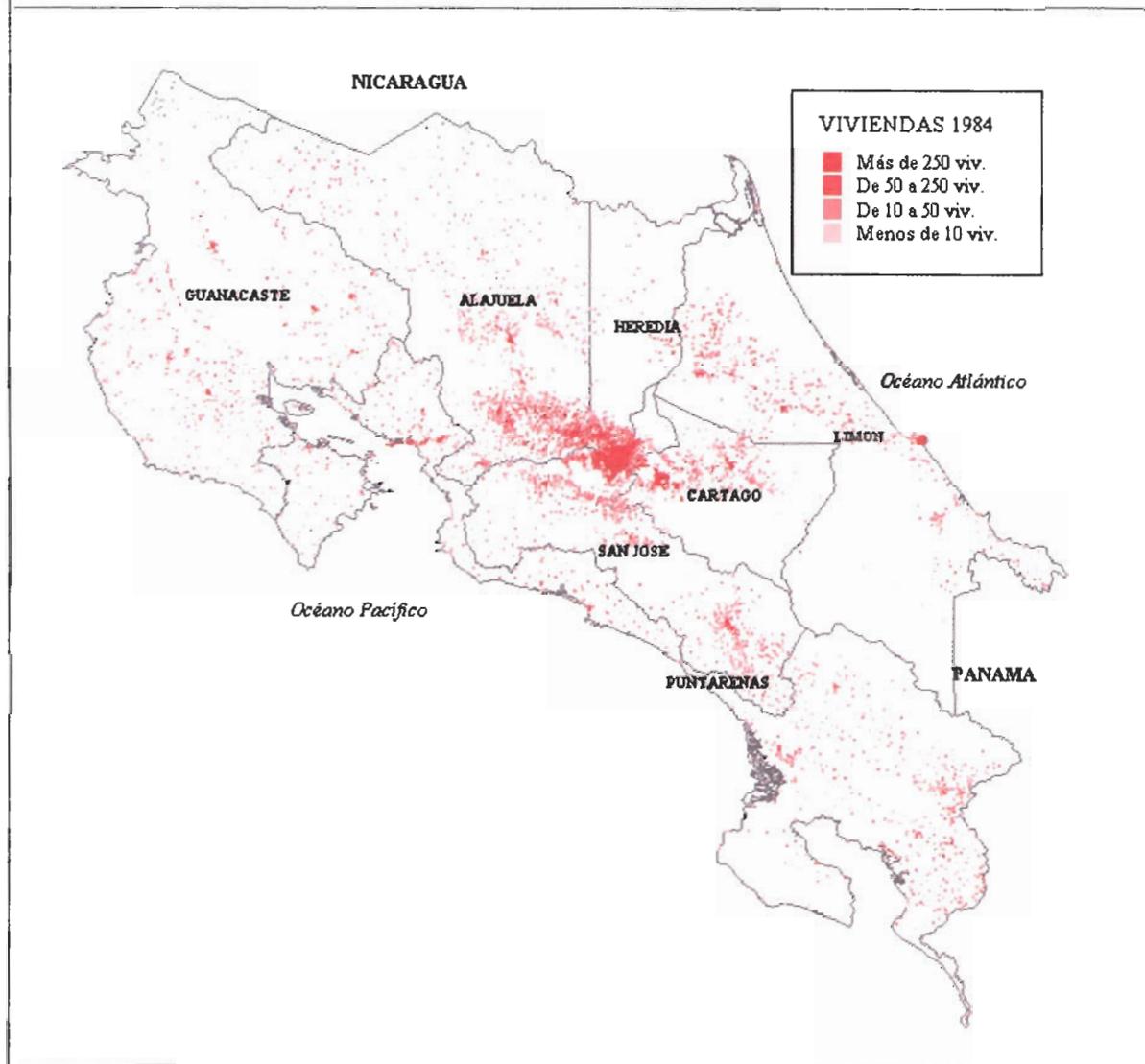


Figura 6

Costa Rica: mapa del número de viviendas rasterizado, Censo 2000.

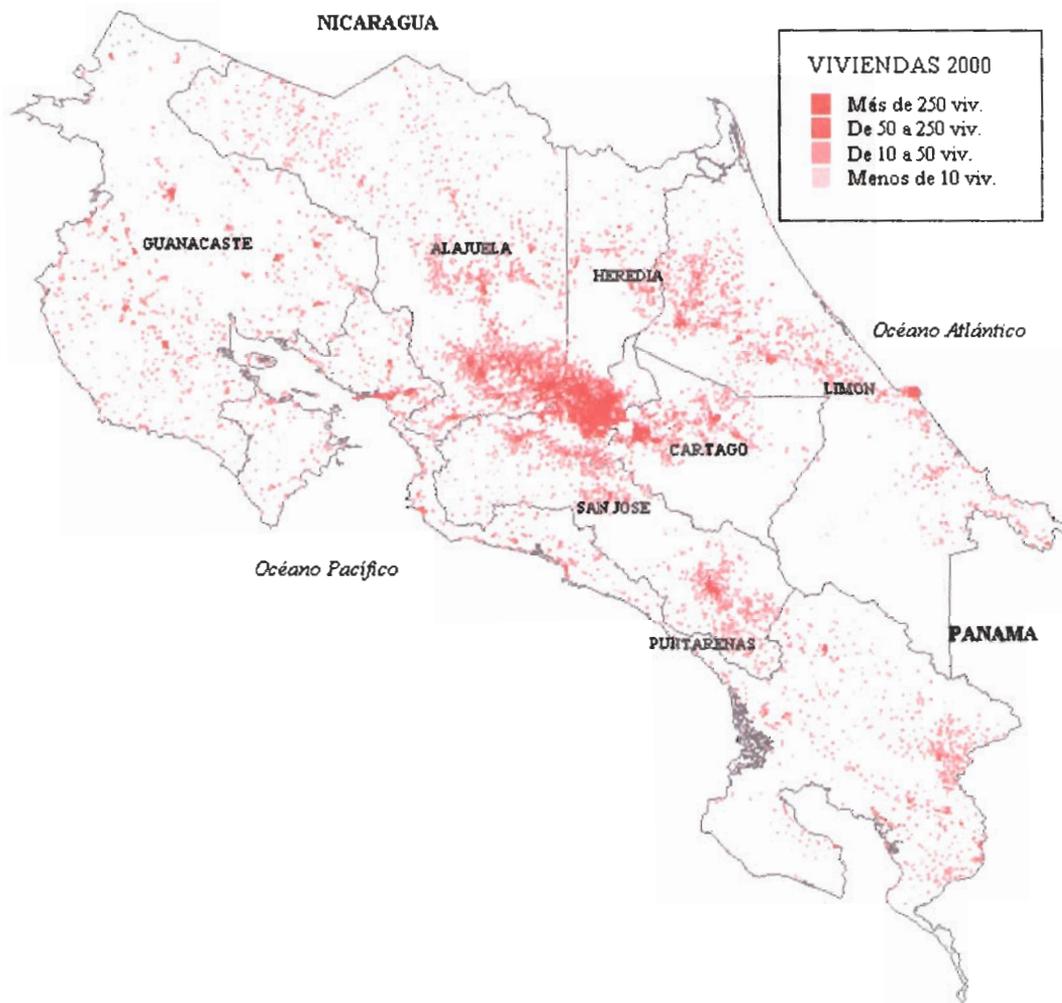


Figura 7

Costa Rica: mapa de la población rasterizado, Censo 1973.

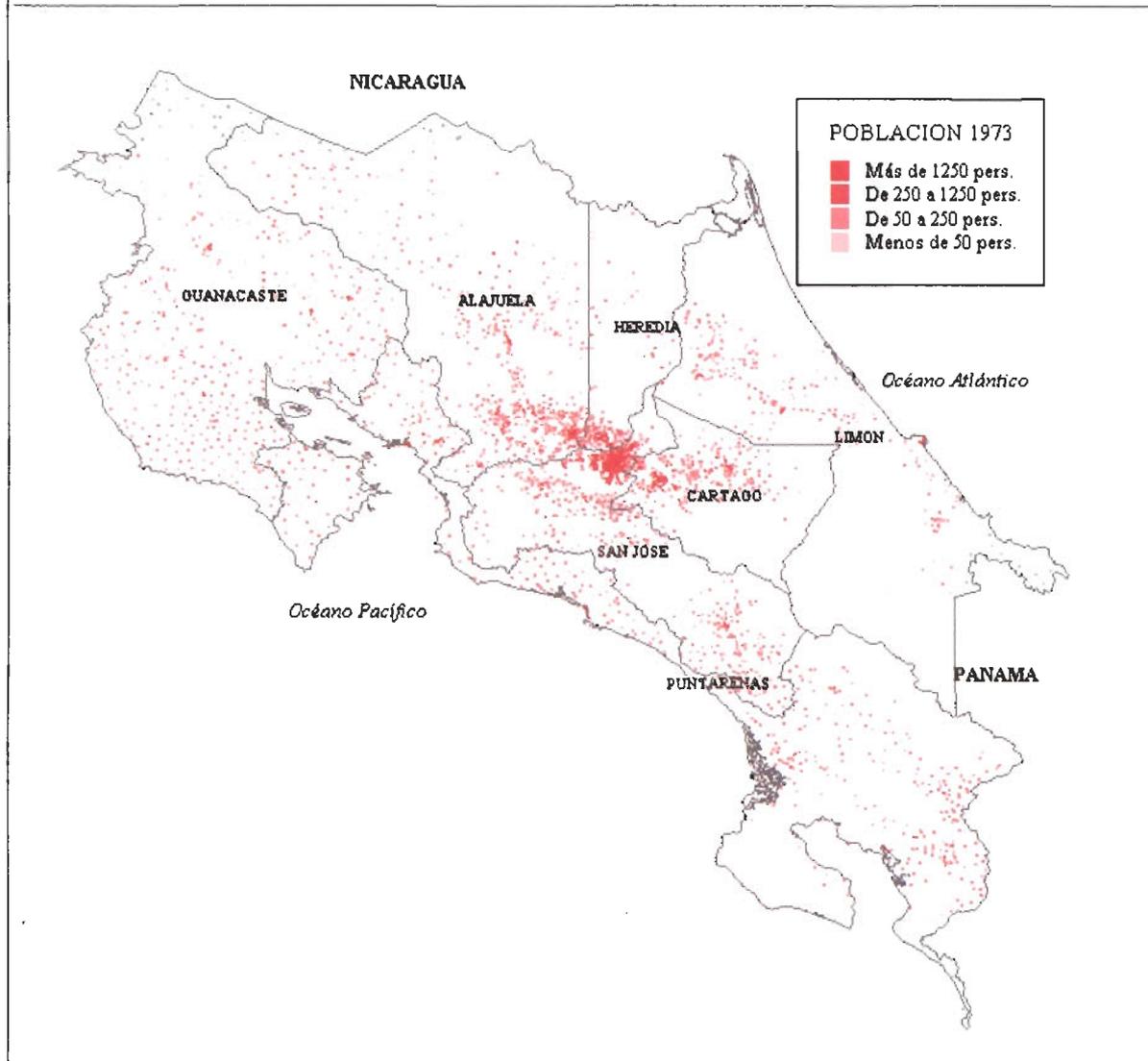


Figura 8

Costa Rica: mapa de la población rasterizado, Censo 1984.

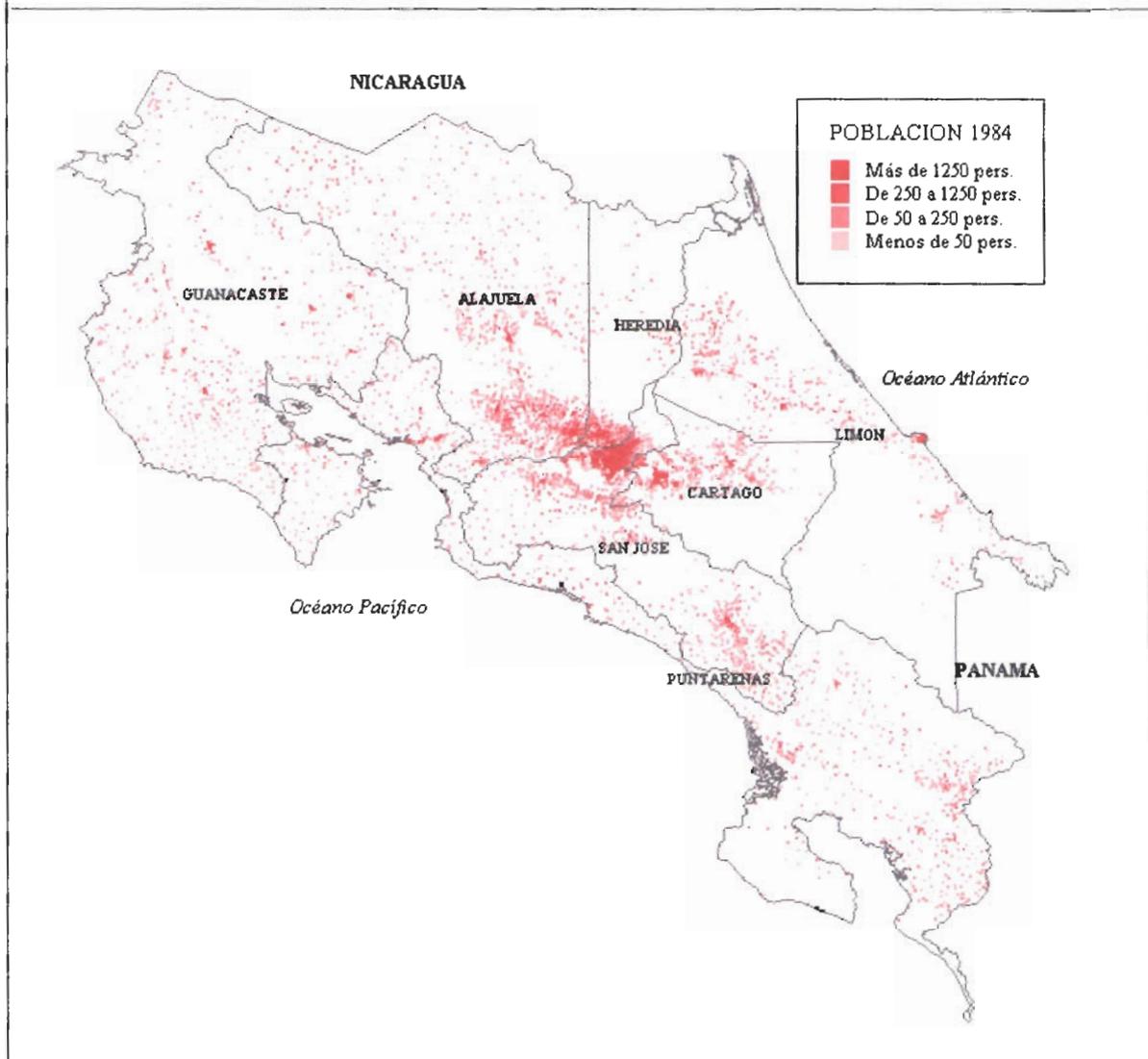
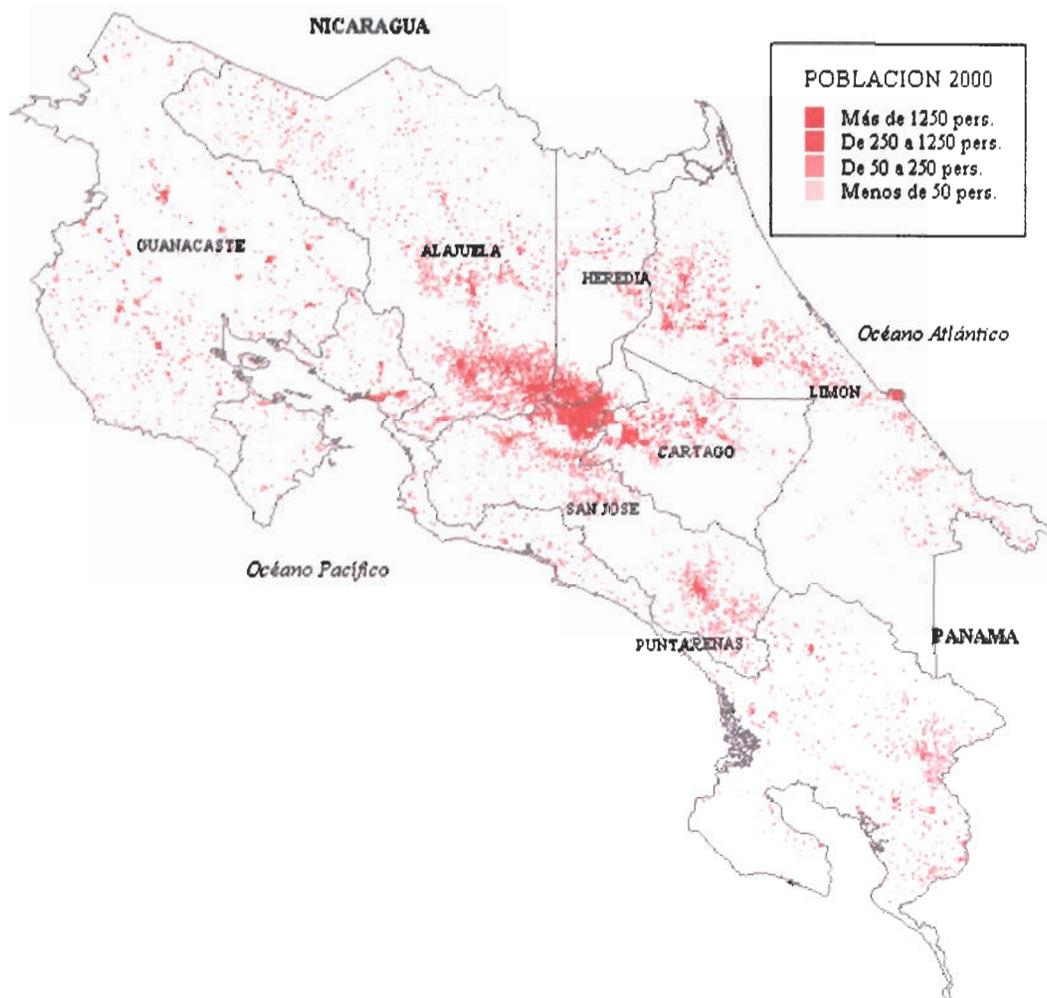


Figura 9

Costa Rica: mapa de la población rasterizado, Censo 2000.



AGRADECIMIENTOS

Los autores agradecen al Instituto Nacional de Estadística y Censos (INEC) por facilitar las bases de datos de los censos de población y vivienda 1973, 1984 y 2000, la cartografía censal y la bitácora de los distritos Puerto Jiménez y Sierpe (rural). Douglas Güell realizó la geocodificación de las casas del distrito rural utilizado. El Centro Centroamericano de Población (CCP) efectuó la geocodificación del distrito urbano de San Pablo.

BIBLIOGRAFÍA

- Bracken, I. (1989). The Generation of socioeconomic surfaces for public policy making. *Environment and Planning A* 21, (4): 307-325.
- Bracken, I. y D. Martin. (1989). The Generation of Spatial Population Distributions from Census Centroid Data. *Environment and Planning A*, 21 (4): 537-43.
- Bracken, I. y D. Martin. (1995). Linkage of the 1981 and 1991 Censuses using Surface Modeling Concepts. *Environment and Planning A*, 27: 379-90.
- Goodchild, M. F., L. Anselin y U. Deichmann. (1993). A Framework for the Areal Interpolation of Socioeconomic Data. *Environment and Planning A*, 25: 383-397.
- Johnson, D. (2000). *Métodos Multivariados aplicados al Análisis de Datos*. Internacional Thomson Editores. México DF, México.
- Martin, D. (1999). SurPop V.2.0: *Background to Census Surface Models*. <http://www.census.ac.uk/cdu/software/surpop/background.html>
- Nordbeck, S. y B. Rystedt. (1970). *Isarithmic Maps and the Continuity of Reference Interval Functions*. *Geografiska Annaler*, 52B: 92-123.
- Rosero-Bixby, L. y A. Palloni. (1998). Population and deforestation in Costa Rica. *Population and Environment: A Journal of Interdisc. Studies*, 20: 149-185.
- Schmid, C. F. y E. H. MacCannell. (1955). Basic Problems, Techniques and Theory of Isopleth Mapping. *Journal of the American Statistical Association*, 50: 220-239.
- Tobler, W. R. (1979). Smooth Pycnophylactic Interpolation for Geographical Regions. *Journal of the American Statistical Association*, 74: 519-530.
- Unisys Corporation. 1985-1998. *MapInfo Professional Version 5.0*. Welch Licensing Department, PA. Estados Unidos.