

Modelización cartográfica de densidades mediante estimadores Kernel¹

Antonio MORENO JIMÉNEZ
Departamento de Geografía
Universidad Autónoma de Madrid

1. Introducción

La presentación gráfica de datos hace posible, a la par que una mejor intelección de muchos fenómenos, una profundización en la información que poseen. No es, por tanto, sólo cuestión estética. Mediante las técnicas oportunas pueden desvelarse pautas más generales y esquemáticas de una forma harto intuitiva y asequible. En Geografía, la necesidad de expresar distribuciones de hechos en el espacio ha conducido al uso de un buen número de métodos que facilitan la visualización de aquéllas. Uno de los aspectos que constituye una constante en el análisis geográfico son las densidades; ya de población, de resurgencias o de cualquier otro suceso o hecho, la cartografía de densidades ha sido un recurso habitual en nuestra disciplina.

En este trabajo se pretende presentar una técnica prácticamente inédita hasta ahora entre los geógrafos² aplicable a la presentación cartográfica de densidades. Este tipo de mapas cuenta con bastante tradición y uso por su significación y expresividad. Habitualmente, su elaboración se sustenta sobre el cociente entre dos tipos de valores: la del fenómeno a estudiar (población por ejemplo) y la superficie de referencia (tamaño de hectáreas o km cuadrados). De esta forma se consiguen mapas en los que deben existir unas unidades espaciales discretas, a las cuales referir las densidades. Los problemas e inconvenientes de la delimitación de estas unidades son importantes y ya bastante conocidos. La metodología que presentamos supera tales problemas al prescindir de dichos ámbitos discretos, que condicionan, y con frecuencia distorsionan fuertemente, el resultado gráfico. Más concretamente, la técnica en cuestión resulta idónea siempre que se cuente con series de puntos localizados mediante sus coordenadas X e Y. Es decir, los datos necesarios son pares de abcisas y ordenadas para el conjunto de los hechos a estudiar.

El tipo de problema que puede resolverse es el de obtener, a partir de esa nube de puntos localizados, el esquema de densidad que conforman. Se trata por tanto de alcanzar una representación de la «tendencia» o pauta global de su distribución, dónde existe más concentración, dónde menos, con qué gradiente se produce la variación espacial, etc. El espacio está presente, pero no bajo la forma de compartimentos estancos más o menos artificiales. El «producto» gráfico final es sencillamente un mapa de isopletas o isodensidades, pero la originalidad y novedad de este trabajo no estriba en el aspecto de dichos mapas que siguen las pautas clásicas, sino en el método de cómputo de los valores de densidad a representar. En los dos apartados siguientes se describen matemática e intuitivamente la naturaleza de los estimadores Kernel y los elementos que quedan al arbitrio del investigador. Seguidamente, y recurriendo a dos conjuntos de datos empíricos, se presentan y valoran los resultados obtenibles mediante una rica gama de mapas.

2. La formulación de los estimadores Kernel

Propuestos por primera vez por Rosenblatt (1956), la singularidad de estos estimadores radica en que adoptan un enfoque no paramétrico para la formulación de funciones de densidad. Con ello no se imponen asunciones tan rígidas a los datos como sucede en el planteamiento paramétrico. Según señala Silverman (1986, p. 1), asumiendo que la distribución de datos posee una densidad de probabilidad f , se les posibilita «hablar» más por sí mismos al determinar dicha función f , que si se exigiera su adecuación a una determinada familia de funciones paramétricas.

Desde el punto de vista matemático estos estimadores poseen un cierto número de propiedades interesantes:

- a) Son funciones simétricas centradas en cero y su integral vale la unidad.
- b) Puesto que el Kernel es una función de densidad, el estimador \hat{f} también lo es.
- c) En la mayoría de los casos (dependiendo del Kernel) se trata también de funciones continuas y diferenciables.
- d) Los Kernels, en la definición que aquí usaremos, son positivos.

Genéricamente, y para el caso univariado, dicho estimador puede escribirse así:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K \left(\frac{x-x_i}{h} \right)$$

siendo x = punto en el cual se trata de estimar la densidad,

x_i = valor de la variable en el caso $i=1, \dots, n$,

K = símbolo del Kernel,

h = anchura de la ventana o parámetro de suavizado. Como veremos más adelante, tal parámetro resulta crucial en este método.

No existe un solo tipo de Kernel. La producción escrita ha ido acumulando un cierto número de ellos de los cuales algunos reproducimos en el cuadro 1.

Cuadro 1(*)
Tipos de Kernels

KERNEL	K(u)
Uniforme	$\frac{1}{2}$, para $ u \leq 1$
Triangular	$(1- u)$, para $ u \leq 1$
Epanechnikov	$\frac{3}{4}(1-u^2)$, para $ u \leq 1$
Cuártico o Bi-weight de Tukey	$\frac{15}{16}(1-u^2)^2$, para $ u \leq 1$
Triweight	$\frac{35}{32}(1-u^2)^3$, para $ u \leq 1$
Gaussiano	$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right)$
Coseno	$\frac{\pi}{4} \cos\left(\frac{\pi}{2}u\right)$, para $ u \leq 1$
K ₂ de Silverman	$\frac{3}{\pi}(1-u^2)^2$, para $ u \leq 1$

Donde $u=(x-h)/h$.

(*) Reproducido parcialmente de Härdle (1991, p. 45). La formulación del Kernel K₂ es la ofrecida por Silverman (1986, p. 767).

Intuitivamente expresado, el Kernel sobre un punto x tomaría una determinada forma según el tipo del mismo. En las Figuras 1, 2 y 3 se presentan algunos de los Kernels detallados en el cuadro anterior. Naturalmente la obtención de la función de estimación puede realizarse en puntos diferentes a los datos observados. La idea básica consiste, en conclusión, en calcular para unos determinados puntos, la suma promediada (de ahí que el estimador suponga sumar sobre n y dividir luego por este valor) de los Kernels o «elevaciones» centradas sobre las observaciones. La Figura 4 ilustra muy gráficamente la técnica para el caso del Kernel gaussiano. El tipo de Kernel determina la forma de la «elevación» y el parámetro h , su anchura. La Figura 5 ejemplifica el efecto de un valor diferente, más pequeño, del parámetro de suavizado h . Compárese con la Figura 4.

La función de estimación Kernel multivariada no sería más que una extensión de las formulaciones anteriores. De esta manera, y definiendo el vector

$$\tilde{x} = (x_1, x_2, \dots, x_d)$$

como el punto x , cuya densidad se trata de estimar, en cada una de las d dimensiones (2 en el caso de las coordenadas geográficas), el valor de la función de densidad para dicho punto se escribiría

$$\hat{f}(\tilde{x}) = \frac{1}{n} \sum_{i=1}^n \left(\prod_{j=1}^d h_j^{-1} \right) K \left(\frac{\tilde{x}_1 - X_{i1}}{h_1}, \dots, \frac{\tilde{x}_d - X_{id}}{h_d} \right)$$

Dicha expresión viene a significar que la estimación para cada punto viene dada por el producto de los Kernels unidimensionales marginales; para el caso del plano tendríamos dos dimensiones o variables (abscisa y ordenada) y por ende, el valor del estimador sería el producto de los Kernels simples en X e Y dividido por las anchuras de ventana respectivas. Naturalmente habría que sumar sobre la totalidad de los datos y dividir por n .

3. La discrecionalidad en la aplicación de estimadores Kernel: el tipo de Kernel y el tamaño de la ventana

Dos elementos quedan sujetos a la discrecionalidad del estudio a la hora de aplicar esta metodología, por un lado el tipo de Kernel y por otro la magnitud del parámetro de ventana. Lo consideraremos brevemente.

Dada la variedad de Kernels es obvio que la formulación adoptada condicione el resultado obtenido. Como se verificará en los experimentos que se presentarán, ello es cierto, si bien salvo algún Kernel concreto, la mayoría de los aquí ensayados, ofrecen una similitud importante en los mapas finales. Tanto Silverman (1986, p. 43) como Härdle (1991, p. 78) coinciden al afirmar que la elección del Kernel es casi irrelevante de cara a la «bondad» de la estimación. La pérdida de calidad de ciertos Kernels respecto al óptimo ha sido evaluada y, para algunos de los aquí estudiados, puede decirse que esa degradación ronda

Figura 1

Kernel uniforme (puntos), triangular (línea continua) y de Epanechnikov (línea discontinua), según Härdle (1991, p. 46)

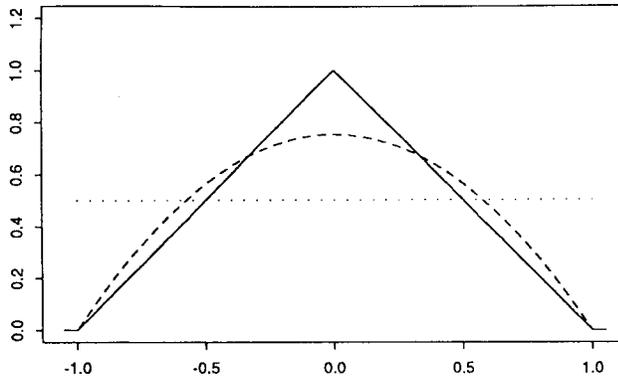


Figura 2

Kernel coseno (puntos), cuártico (línea continua) y Triweight (línea discontinua), según Härdle (1991, p. 46)

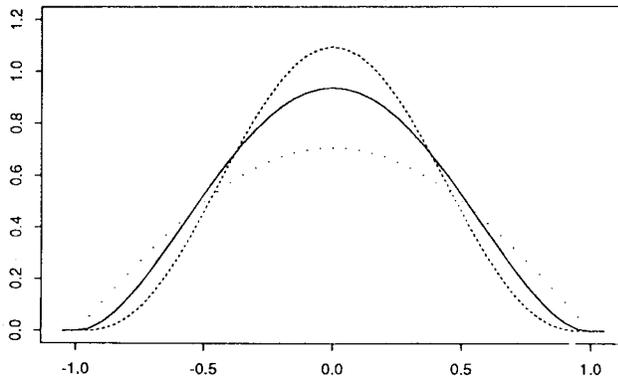


Figura 3

Kernel gaussiano, según Härdle (1991, p. 47)

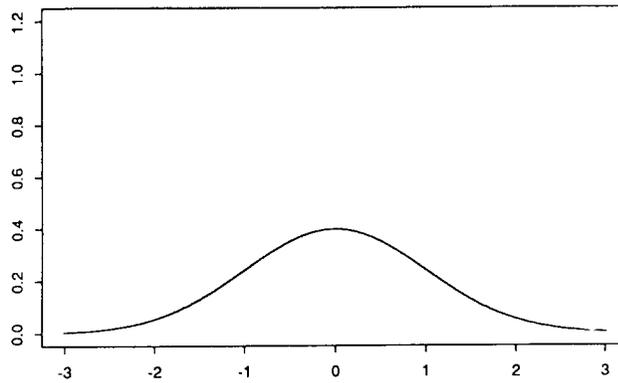


Figura 4
Estimación con un Kernel gaussiano mostrando los Kernels individuales con $h=0,4$, según Silverman (1986, p. 14)

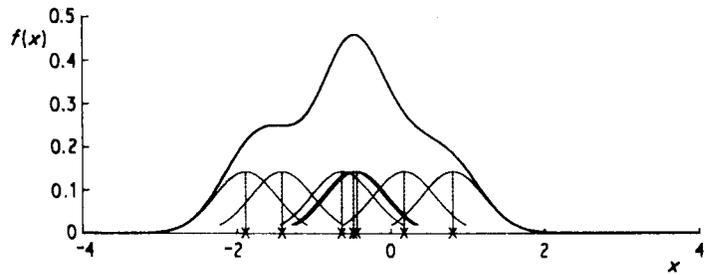
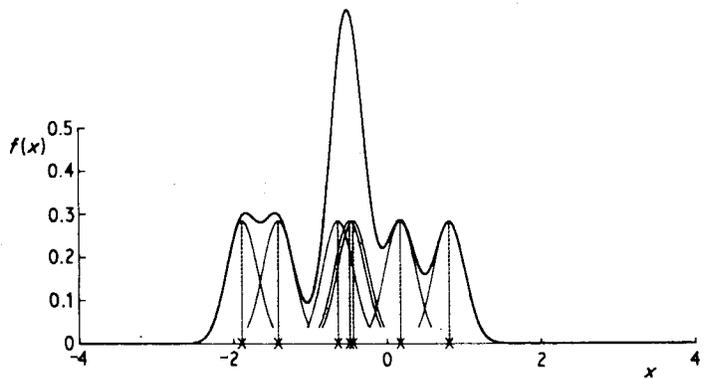


Figura 5
Estimación con un Kernel gaussiano mostrando los kernels inividuales con $h=0,2$, según Silverman (1986, p. 14)



el 6 por ciento en el caso del Kernel Uniforme y Triweight, y el 4 por ciento para el Gaussiano. El Triangular Cuártico o Coseno está en torno al 1 por ciento o menos. Por tanto quizás otras consideraciones, como por ejemplo los costes de cálculo, deben guiar la elección.

Mucho más crítico que el tipo de Kernel resulta el tamaño de la ventana. Su elección puede realizarse según tres enfoques:

a) Una elección subjetiva, a partir de ensayos con diversos valores de h . Como aquí se verá, la representación cartográfica de los resultados obtenidos en cada caso puede constituir una base razonable para ello, permitiendo elegir aquella que consigue expresar mejor el mensaje deseado por el analista. El *trade-off* entre la «rugosidad» de los mapas resultantes con ventanas pequeñas y el hipersuavizado que generan las grandes puede y debe ser juiciosamente resuelto.

b) Si es posible asumir que una distribución conocida subyace a los datos es posible fundar la elección sobre tal base. El criterio sería entonces el de elegir aquel tamaño de ventana que minimiza el llamado MISE (Mean Integrated Square Error), traducible como la media de los errores cuadráticos entre los valores obtenidos por nuestra función de estimación Kernel y los de la verdadera función de densidad (generalmente desconocida). Para el caso de que sea asumible que una distribución normal es la verdadera han sido determinadas fórmulas aproximadas para la elección de la ventana que usan la desviación típica (α) o, una medida más robusta, la amplitud intercuartil (R). Su expresión concreta sería:

$$h=A\sigma n^{-0.2}$$

o bien

$$h=B R n^{-0.2}$$

las constantes A y B las ha resumido Härdle (1991, p. 228) en el cuadro 2.

c) Métodos automáticos que tratan de obtener el tamaño de ventana que minimiza alguna medida expresiva de la diferencia entre la función de densidad verdadera y el estimador utilizado. Su complejidad aconseja omitir por ahora una exposición más detallada de ellos (cf. Härdle, 1991, cap. 4 y Bowman, 1985).

4. Aplicación a la representación de áreas de atracción de equipamientos

Uno de los posibles campos de aplicación de la cartografía de densidades estriba en la representación de la demanda espacial de servicios: la nube de puntos que marcan el domicilio u origen del desplazamiento de los clientes de comercios, de equipamientos co-

Cuadro 2
Parámetros para fórmulas de ventana

KERNEL	A	B
Uniforme	1.84	1.37
Triangular	2.57	1.91
Epanechnikov	2.35	1.75
Cuártico o Bi-weight	2.78	2.07
Triweight	3.16	2.35
Gaussiano	1.06	0.79
Coseno	2.43	1.81

lectivos u otros servicios puede ser objeto de tratamiento con estimadores Kernel para obtener una imagen que exprese la densidad de ellos y su variación en el espacio. Dos ejemplos reales han sido utilizados aquí: por un lado los usuarios de un Centro de la Tercera Edad, del barrio de Lavapiés y por otro los del Centro Cultural Nicolás Salmerón, del distrito de Chamartín, ambos en Madrid. Los datos proceden de encuestas realizadas entre los visitantes a los mismos que han sido analizadas en otro lugar (cf. Jiménez *et alii*, 1989 y Moreno *et alii*, en prensa). En ambos casos no se han incluido la totalidad de los datos, sino sólo la fracción de usuarios que procedían de las inmediaciones. Tal reducción en la información no es gratuita. En las pruebas preliminares se constató que, dada la existencia de algunos visitantes de zonas bastante distantes, ello ocasionaba representaciones gráficas poco detalladas y, por ende escasamente susceptibles de cotejo y valoración. En las Figuras 6 y 7 se pueden observar la disposición de la nube de visitantes procedentes de la zona próxima a ambos equipamientos.

Con tal fin en este trabajo se han utilizado dos programas (en FORTRAN) que operacionalizan los estimadores Kernel presentados en el cuadro 1. Uno de ellos (programa KPLOT³) original de Donthu y Rust (1989), y otro (KERNEL) elaborado por el autor. El primero de ellos resuelve el Kernel de tipo Gaussiano, y el otro los restantes. En síntesis, ambos obtienen, a partir de la lista de pares de coordenadas de los puntos, la estimación de densidad en una malla cuadrada. La representación gráfica ulterior ha sido realizada mediante el sistema SURFER, con el cual es compatible el formato de datos de la malla.

Con objeto de examinar la sensibilidad de los resultados a los dos ítems susceptibles de manipulación por el usuario se han realizado diversas pruebas, alterando el tipo de Kernel y la anchura de la ventana, que seguidamente se presentan.

4.1. Estudio experimental del efecto del tipo de Kernel

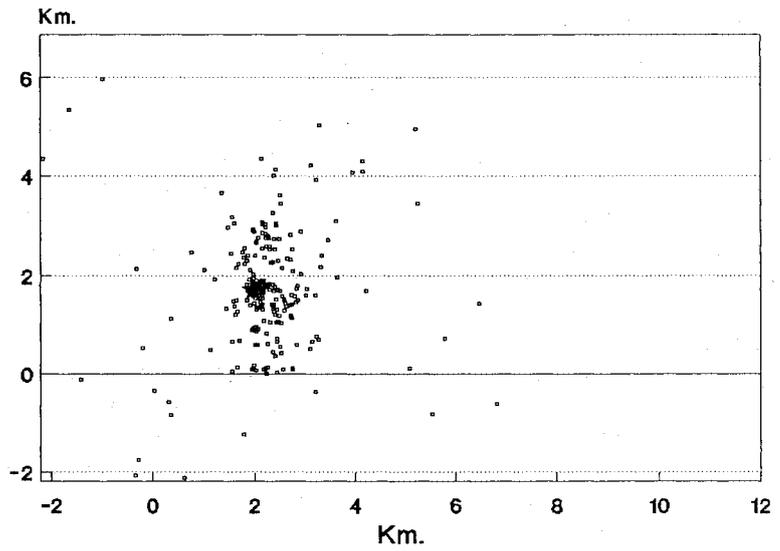
A la hora de aplicar cada tipo de Kernel a los datos anteriores fue necesario adoptar una ventana concreta. El criterio seguido en cada caso fue como sigue. La distribución especial de los visitantes a un punto de servicio puede ser muy variable en función de factores diversos, pero lo que se constata rápidamente es la fuerte concentración en las proximidades del mismo. Tal circunstancia hace que los datos recuerden a menudo a una distribución bivariada normal o análoga (Student por ejemplo). Por tal motivo la elección de la ventana (h_x y h_y) se realizó en cada caso de acuerdo con la fórmula dada en el apartado 3 (*supra*) con la desviación típica.

Los resultados (figuras 8 y 9) muestran una concordancia fundamental en los mapas derivados de los diferentes Kernels, reflejando claramente la fuerte densidad en torno a los centros de servicios⁴ y el rápido descenso con la distancia. Las formas más discrepantes corresponden al Kernel uniforme que produce estimaciones de las variaciones espaciales excesivamente escalonadas y poco verosímiles. Salvo este caso, se puede corroborar la premonición hecha antes de que el tipo de Kernel es relativamente poco relevante, por la similitud de sus resultados.

4.2. Estudio experimental del efecto de anchura de la ventana

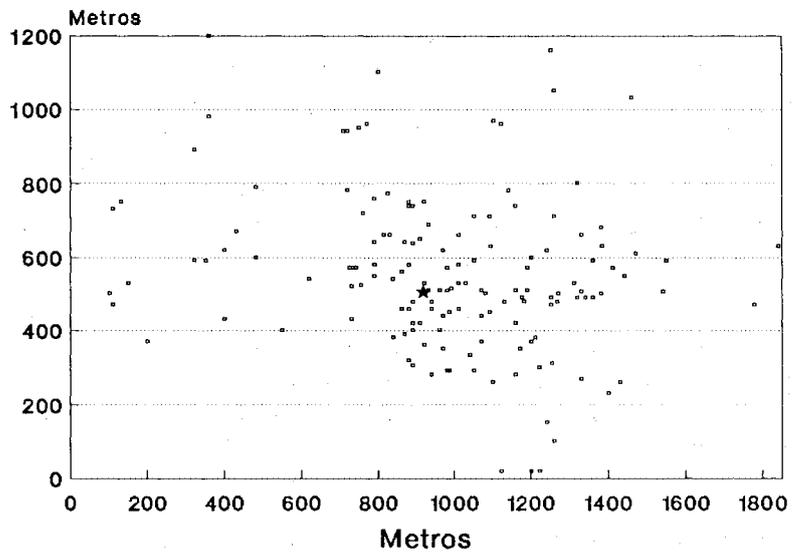
El efecto más espectacular de la metodología que estamos analizando se produce cuando se modifica el tamaño de la ventana de suavizado. Para tal fin, y utilizando siempre

Figura 6
Distribución espacial de los visitantes al Centro Cultural de Chamartín (Madrid)



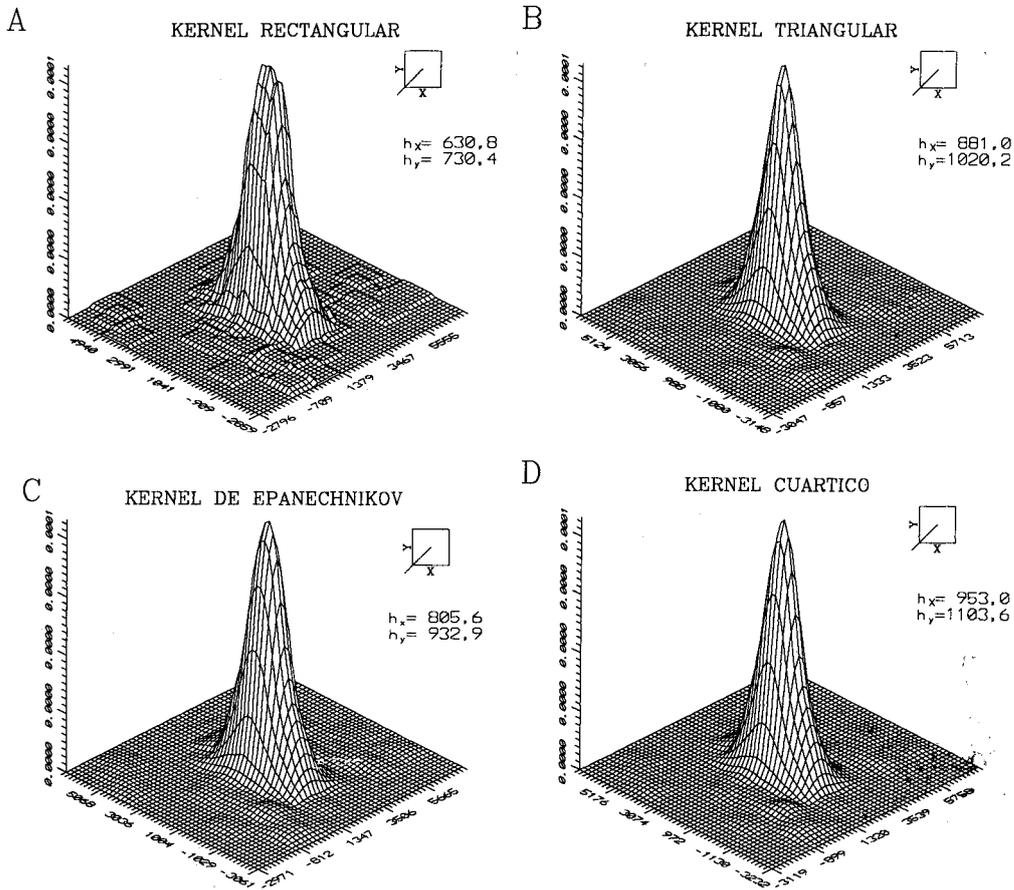
(Zona próxima)

Figura 7
Distribución espacial de los visitantes al Centro de la Tercera Edad de Lavapiés (Madrid)



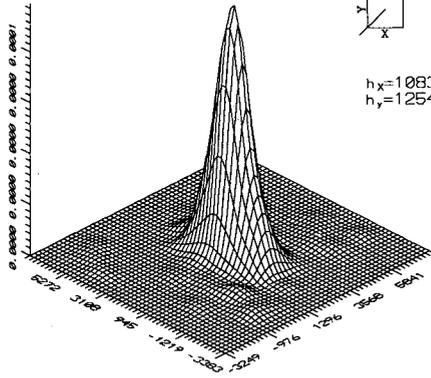
(Zona próxima)

Figura 8 A-H
Mapas de densidad de usuarios al Centro Cultural según los diversos estimadores Kernel



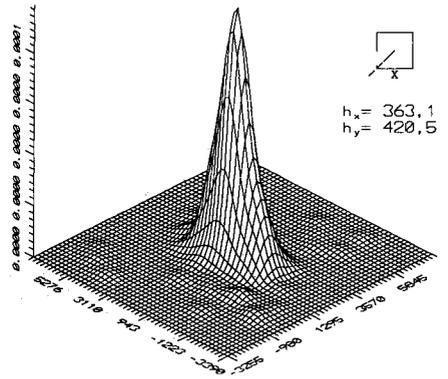
E

KERNEL TRIWEIGHT



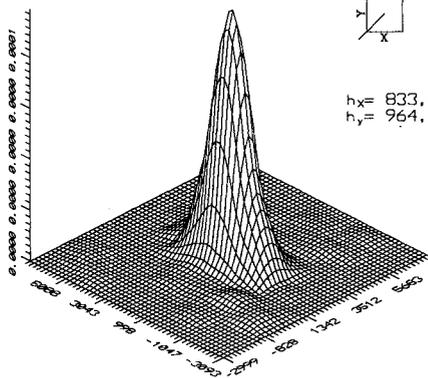
F

KERNEL GAUSSIANO



G

KERNEL COSENO



H

KERNEL K(2) DE SILVERMAN

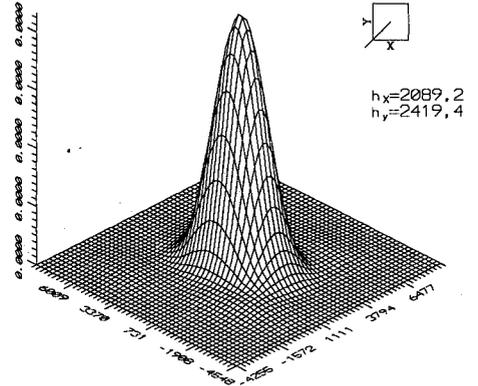
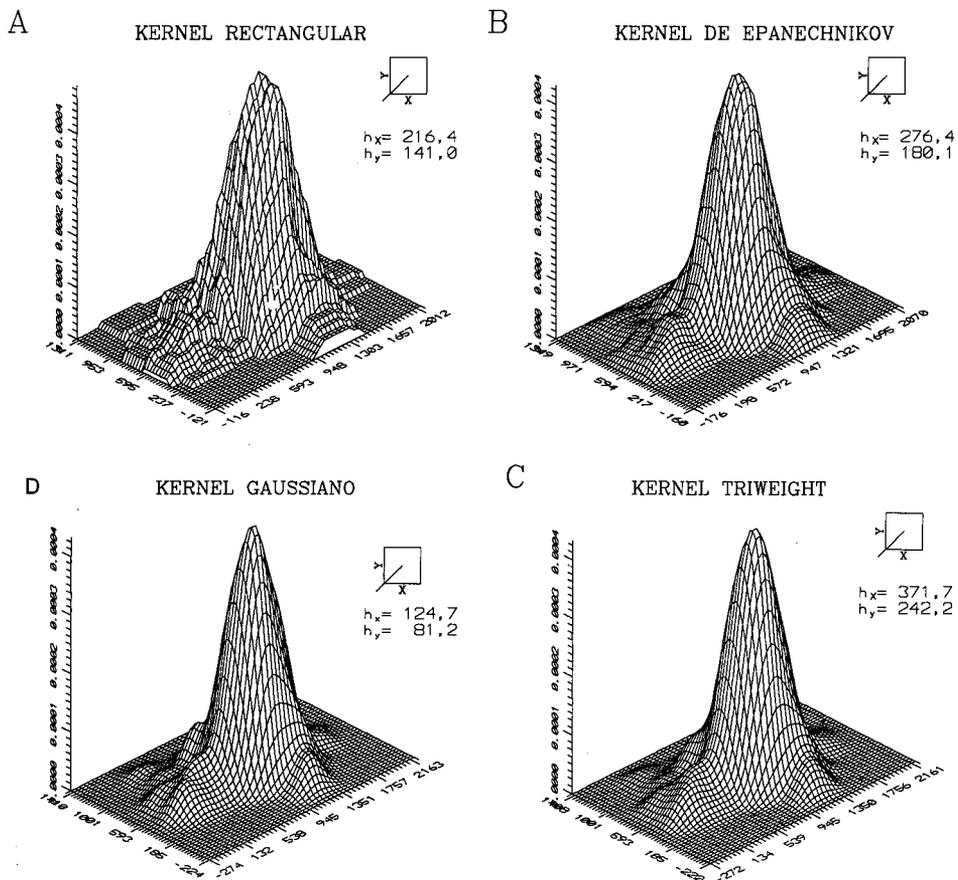


Figura 9 A-D
Mapas de densidad de usuarios al Centro de la Tercera Edad según algunos
estimadores Kernel



el mismo Kernel, el Coseno, se han aplicado diferentes valores del parámetro h a los datos del Centro Cultural. Los resultados pueden seguirse en la Figura 10. La conclusión básica es palmaria: a medida que aumentamos la ventana se provoca un suavizado creciente en el mapa resultante. Valores pequeños de la misma hacen aflorar paulatinamente la «rugosidad» de los datos originales. A partir de las figuras queda en evidencia la necesidad de realizar una elección meditada y justificable de dicho parámetro, en función de los fines del estudio.

4.3. Evaluación

Dos cuestiones finales deben ser traídas a colación. Por un lado, los costes de obtención de los mapas y por otro su fiabilidad. En lo concerniente al primer punto, el principal elemento a considerar son los costes de cómputo. El algoritmo adoptado por el programa KERNEL optimiza los cálculos de manera que, para cada dato sólo se requiere evaluar un limitado número de puntos en la malla resultante (aquellos para los que la distancia entre ambos es inferior a la anchura de la ventana), ya que en virtud de la fórmula de los Kernels cuando tal distancia es mayor su efecto se anula. Por tal motivo la mayoría de los ejemplos aquí presentados se obtienen en un lapso breve (uno o dos minutos con un procesador 80386 SX a 16 Mh con coprocesador). Un caso aparte es el del Kernel gaussiano. La no limitación de la influencia de cada dato a un «ámbito», definido por la ventana, obliga a estimar su contribución para la totalidad de los puntos de la malla. Ello significa tiempos de ejecución superiores; así por ejemplo la malla de 50×50 obtenida a partir de 155 datos de usuarios del Centro de la Tercera Edad requirió 21 minutos con un PC 80386 a 20 Mh. Los 238 puntos de los usuarios al Centro Cultural requirieron 3 horas y 9 minutos sobre un PC 8086 a 8 Mh y 43 minutos aproximadamente con un procesador 80386 SX y coprocesador matemático. Al ser la duración total función del número de puntos y del tamaño de la malla deberán tenerse en cuenta ambos.

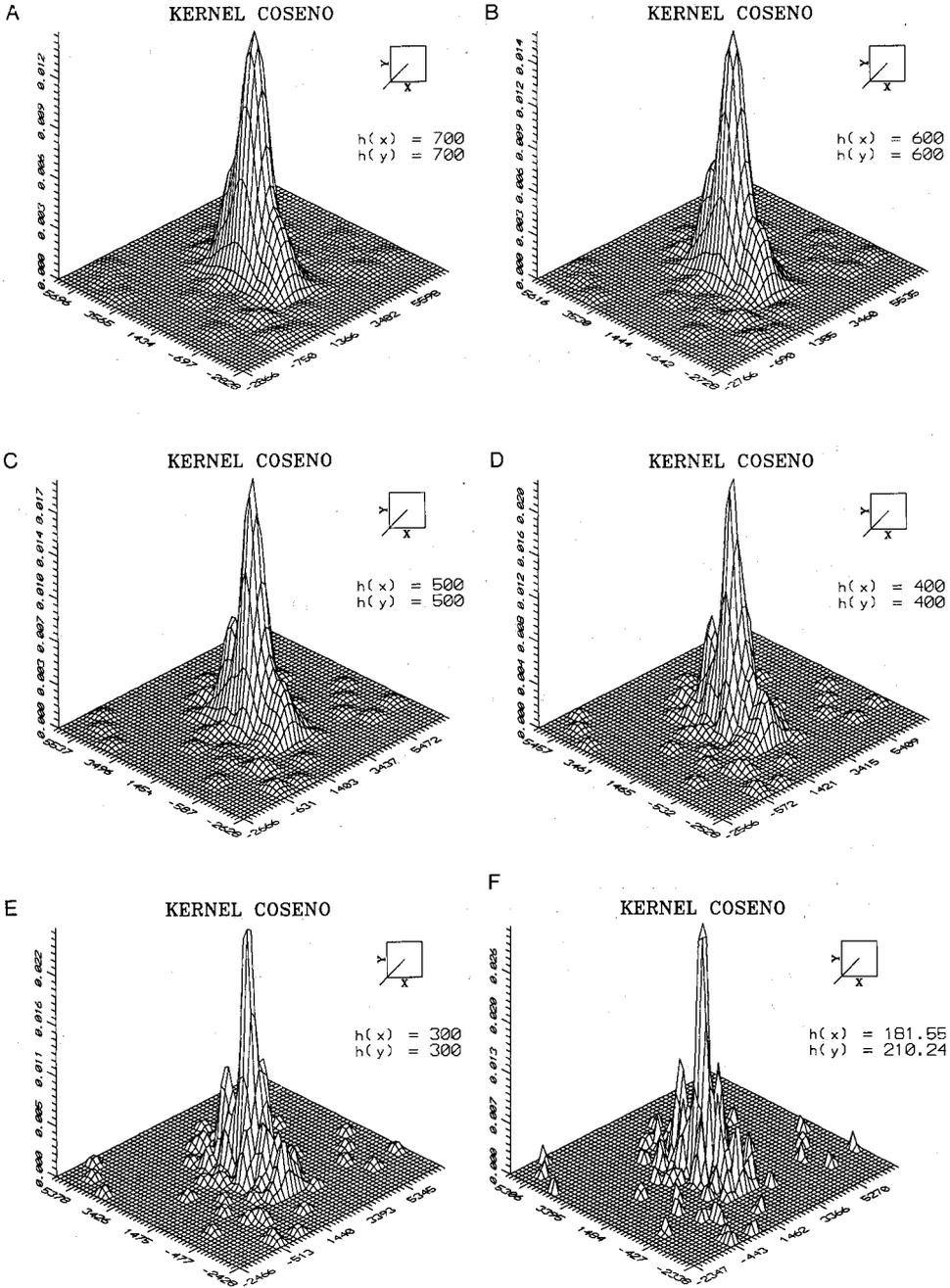
En lo concerniente a la fiabilidad de las estimaciones no se ha realizado por el momento una investigación sistemática. Se puede citar, sin embargo, el trabajo de Donthu y Rust (1989) que, para el caso del Kernel gaussiano, llevaron a cabo una comparación de este método con otros de estimación de densidades (histogramas, *squared surface density analysis* y ajuste de una distribución normal bivariada), llegando a la conclusión de que dicho Kernel era superior a las restantes técnicas, al cotejar sus resultados con los de dichos métodos en un text para intentar replicar (estimar) distribuciones conocidas.

5. Conclusiones

La elaboración de cartografía de densidades es una práctica habitual entre los geógrafos. En el presente trabajo hemos presentado una técnica nueva que permite obtener mapas de isopleas expresando la densidad de ocurrencia de fenómenos en el espacio. La estructura matemática de los estimadores Kernel, en los que se basa esa técnica, permiten obviar uno de los problemas más recurrentes en los mapas de densidades basados en áreas discretas (coropletas). Pese a todo, las densidades obtenidas mediante estimadores Kernel requieren del estudioso una elección del tamaño de la ventana de suavizado que, como recomendación general, debe hacerse de forma congruente con el «mensaje» que se desee transmitir o los aspectos a realzar. La elección del tipo de Kernel parece de importancia secundaria, si exceptuamos el uniforme o rectangular.

Figura 10 A-F

Mapas de densidad de usuarios al Centro Cultural según el Kernel coseno y diferentes tamaños de la ventana



Entre las múltiples aplicaciones posibles que este tipo de mapas puede tener, en este breve trabajo se ha expuesto una concreta, la representación de la distribución espacial de los usuarios de equipamientos colectivos. Se consigue así una imagen mucho más simple y general del área de impacto o beneficiada por esa inversión pública. Cabe asimismo, realizar variantes de dichos mapas seleccionando fracciones de la demanda: grupos según edad, sexo, nivel de instrucción u ocupación pueden ser representados y comparados entre sí, posibilitando valoraciones comparativas de sus respectivas distribuciones. Asimismo la inserción de esta metodología en los sistemas de información geográfica abre perspectivas halagüeñas para aplicaciones diversas que se explorarán en el futuro.

Notas

- ¹ Investigación realizada dentro del proyecto PB88-0181 de la DGICYT.
- ² Como antecedentes sólo caben mencionar el trabajo inédito de Brusndon (1990) y en las ciencias del marketing el estudio de Donthu y Rust (1989).
- ³ Agradezco a los autores la cortés cesión de dicho programa.
- ⁴ El centro de ancianos se sitúa en las coordenadas X=934 e Y=490. El centro cultural en X=2042,2 e Y=1800.

Bibliografía

- BOWMAN, A.W. (1985): «A comparative study of some kernel-based nonparametric density estimators», *J. Statist. Comput. Simul.*, 21, pp. 313-317.
- BRUNSDON, Ch. (1990): «Techniques for point process information», *Workshop on Spatial Analysis with GIS*, Handouts. EGIS'90. Amsterdam. 8 p.
- DONTHU, N. y RUST, R. (1989): «Estimating geographic densities using kernel density estimation», *Marketing Science*, 8, 2, pp. 191-203.
- HÄRDLE, W. (1991): *Smoothing techniques with implementation in S*, Nueva York, Springer Verlag.
- JIMÉNEZ, A. *et alii* (1989): «La demanda espacial para centros de la tercera edad: un ensayo de definición operativa», *Segundas Jornadas sobre población española*. Palma de Mallorca, Universitat de les Illes Balears, pp. 263-272.
- MORENO, A. *et alii* (en prensa): «Los centros culturales en Madrid: un análisis geográfico de la provisión y el uso», *Estudios Geográficos*.
- ROSENBLATT, M. (1956): «Remarks on some nonparametric estimates of a density function», *Annals of Mathematical Statistics* 27, pp. 832-837.
- RUST, R. y BROWN, J.A.N. (1986): «Estimation and comparison of market area densities», *Journals of Retailing*, 62, 4, pp. 410-430.
- SILVERMAN, B.W. (1986): *Density estimation for statistics and data analysis*, Londres, Chapman and Hall.