

## Bandwidth selection for kernel density estimation of forest fires

## Selección de ancho de banda para la estimación de densidad kernel de incendios forestales

José G. Flores-Garnica<sup>1</sup> & Alejandra Macías-Muro<sup>2</sup>

<sup>1</sup>Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias (INIFAP), Campo Experimental Centro Altos de Jalisco. Parque Los Colomos s/n, col. Providencia. C. P. 44660. Guadalajara, Jalisco, México.

<sup>2</sup>Universidad de Guadalajara, Centro Universitario de Ciencias Biológicas y Agropecuarias. Camino Ramón Padilla Sánchez 2100. C. P. 44600. Nextipac, Zapopan, Jalisco, México.

\*Corresponding author: flores.german@gmail.com, tel.: +52 (331) 169 4533

### Abstract

**Introduction.** The mapping of areas with higher forest fire density can be developed through kernel density estimation, which requires the selection of a function and bandwidth ( $h$ ). The  $h$  value, when defined by subjective (visual) processes, will depend on the knowledge and experience of the person making the selection.

**Objective:** To propose a statistical alternative, based on forest fires information (2005-2013) from Jalisco, Mexico, for the selection of  $h$  as support for kernel density estimation.

**Materials and methods:** A total of 13  $h$  values were defined using seven techniques. The  $h$  value was selected using the following statistics: root mean square error, root mean integrated squared error, coefficient of variation and comparative percentage.

**Results and discussion:** The  $h$  values obtained with the techniques analyzed were between 2 550 and 41 906 m. There was great variation in the results; the range between the maximum and the minimum value was 39 356.34 m with an average of  $10\ 936.74 \pm 9\ 955.04$  m. The above implies that there is no single and universal process for all cases. According to the validation criteria, the statistically most adequate  $h$  value is between 5 300 and 5 900 m; the closest result was obtained with the mean random distance technique (5 395 m).

**Conclusion:** It is possible to select  $h$  under a practical statistical perspective, avoiding the use of subjective criteria.

### Resumen

**Introducción.** La cartografía de áreas con mayor densidad de incendios forestales puede desarrollarse a través de la estimación de densidad kernel, la cual requiere la selección de una función y un ancho de banda ( $h$ ). El valor  $h$ , cuando se define mediante procesos subjetivos (visuales), dependerá del conocimiento y experiencia del seleccionador.

**Objetivo:** Proponer una alternativa estadística, basada en información de incendios forestales (2005-2013) del estado de Jalisco, México, para la selección de  $h$  como soporte en la estimación de la densidad kernel.

**Materiales y métodos:** A través de siete técnicas se definieron 13 valores de  $h$ . El valor de  $h$  se seleccionó usando los estadísticos siguientes: raíz del error cuadrático medio, raíz del error medio cuadrático integrado, coeficiente de variación y porcentaje comparativo.

**Resultados y discusión:** Los valores  $h$  obtenidos con las técnicas probadas se encontraron entre 2 550 y 41 906 m. Hubo gran variación en los resultados; el rango entre el valor máximo y el mínimo fue 39 356.34 m con una media de  $10\ 936.74 \pm 9\ 955.04$  m. Lo anterior implica que no existe un proceso único y universal para todos los casos. De acuerdo con los criterios de validación, el valor de  $h$  estadísticamente más adecuado se encuentra entre 5 300 y 5 900 m; el resultado más cercano se obtuvo con la técnica distancia aleatoria media (5 395 m).

**Conclusión:** Es posible seleccionar  $h$  bajo una perspectiva estadística práctica, evitando el uso de criterios subjetivos.

**Keywords:** Fire density; search area; continuous surfaces; mean random distance; mapping of areas.

**Palabras clave:** Densidad de incendios; radio de búsqueda; superficies continuas; distancia aleatoria media; cartografía de áreas.

## Introduction

Forest fires are one of the most important causes of alteration and loss of forest cover; to counteract forest fires it is necessary to implement control and prevention strategies (Thompson, Calkin, Finney, Gebert, & Hand, 2013). However, due to limited resources, the institutions in charge must locate priority attention areas, which can be determined through several criteria. One of those criteria is based on the classification of areas based on forest fire density (Mohammadi, Bavaghar, & Shabaniyan, 2014), for which there are techniques such as spatial interpolation (Vilar, Martín, & Martínez, 2011) and kernel density estimation (Koutsias, Kalabokidis, & Allgöwer, 2004; Kuter, Usul, & Kuter, 2011). The latter is based on a nonparametric process, whose use involves the selection of a function to estimate the density (Silverman, 1986) and find the appropriate site size to determine it. To do this, a circular site is assumed and an appropriate search area (bandwidth  $h$ ) is defined (Turlach, 1999); however, when the forest fire density is defined, the statistical criteria that support the selection of the  $h$  value are not always clear, even in a visual way (de la Riva, Pérez, Lana, & Koutsias, 2004; Koutsias et al., 2004; Krisp, Peters, Murphy, & Fan, 2009). In this sense, the objective of this study was to establish a statistical process defining the bandwidth  $h$  (maximum search area) required for kernel density estimation, in order to develop a replicable and comparable methodology to map priority areas against forest fires.

## Materials and methods

### Study area

The study was carried out with information from Jalisco, located in the western part of Mexico, where between 17 000 and 20 000 ha are burned annually, with an average of 530 fires (Comisión Nacional Forestal [CONAFOR], 2015). To define the bandwidth ( $h$ ), georeferenced information was used on the fires that occurred in the period 2005 to 2013 (Figure 1).

### Kernel density estimation

Forest fires are georeferenced events that occur differentially in a given region, defining spatial variations in their density (number of fires per unit area) (Salvati & Ferrara, 2015). To map the spatial continuity of this density the following methods are used: 1) classical, which is derived in a) spatial autocorrelation analysis (Anselin, Bera, Florax, & Yoon, 1996), b) cluster analysis (point clustering) (Kulldorff & Nagarwalla, 1995) and c) spatial autoregression models (Anselin et al., 1996); and 2) density maps, where the techniques of quadrant analysis and kernel density are

## Introducción

Los incendios forestales son una de las causas más importantes de la alteración y pérdida de la cobertura forestal; para contrarrestarlos es necesario implementar estrategias de prevención y control (Thompson, Calkin, Finney, Gebert, & Hand, 2013). No obstante, debido a los recursos limitados, las instituciones encargadas deben ubicar áreas de atención prioritaria, las cuales pueden determinarse a través de varios criterios. Uno de estos se basa en la clasificación de áreas con base en la densidad de incendios forestales (Mohammadi, Bavaghar, & Shabaniyan, 2014), para lo cual existen técnicas como la interpolación espacial (Vilar, Martín, & Martínez, 2011) y la estimación de densidad kernel (Koutsias, Kalabokidis, & Allgöwer, 2004; Kuter, Usul, & Kuter, 2011). Esta última se basa en un proceso no paramétrico, cuyo uso implica la selección de una función para estimar la densidad (Silverman, 1986) y encontrar el tamaño de sitio adecuado para determinarla. Para ello, se asume un sitio circular y se define un radio de búsqueda (ancho de banda  $h$ ) apropiado (Turlach, 1999); sin embargo, cuando la densidad de incendios forestales se define, los criterios estadísticos que soportan la selección del valor de  $h$  no siempre son claros, llegando incluso a hacerse simplemente de forma visual (de la Riva, Pérez, Lana, & Koutsias, 2004; Koutsias et al., 2004; Krisp, Peters, Murphy, & Fan, 2009). En este sentido, el objetivo del trabajo fue establecer un proceso estadístico que defina el ancho de banda  $h$  (radio máximo de búsqueda) requerido en la estimación de la densidad kernel, con el fin de desarrollar una metodología replicable y comparable para la generación de mapas de áreas prioritarias contra incendios forestales.

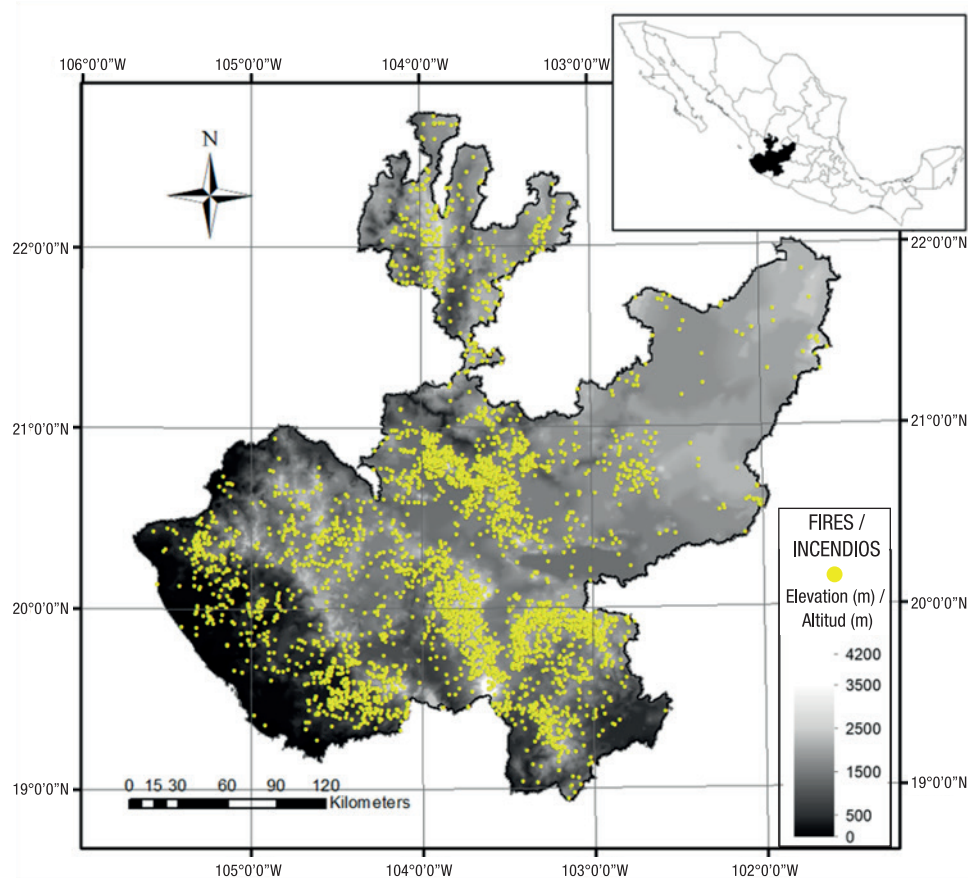
## Materiales y métodos

### Área de estudio

El estudio se llevó cabo con información del estado de Jalisco, ubicado en el occidente de México, donde se queman entre 17 000 y 20 000 ha anuales, con un promedio de 530 incendios (Comisión Nacional Forestal [CONAFOR], 2015). Para definir el ancho de banda ( $h$ ), se usó información georreferenciada de los incendios ocurridos en el periodo 2005 al 2013 (Figura 1).

### Estimación de la densidad kernel

Los incendios forestales son eventos georreferenciados que ocurren en forma diferencial en una región dada, definiendo variaciones espaciales en su densidad (número de incendios por unidad de superficie) (Salvati & Ferrara, 2015). Para cartografiar la continuidad espacial de esta densidad se usan los siguientes métodos: 1) clásico, que se deriva en a)



**Figure 1. Forest fires occurred in Jalisco in the period 2005-2013 (CONAFOR, 2015). Geographical Coordinate System: GCS\_WGS\_1984; Datum: WGS84.**

**Figura 1. Incendios forestales ocurridos en el estado de Jalisco en el periodo 2005-2013 (CONAFOR, 2015). Sistema de Coordenadas Geográficas: GCS\_WGS\_1984; Datum: WGS84.**

used (Fuenzalida, Cobs, & Guerrero, 2013). The latter has been used when historical fire data is available (Koutsias et al., 2004, Kuter et al., 2011). Kernel density is a non-parametric technique based on several functions, some of these are: the square function (Silverman, 1986), the uniform function, the Epanechnikov function, normal distribution, triangular function and quartic function (Turlach, 1999). In this way, the kernel density estimator, for a multivariate case, is defined based on the following mathematical expression (Amatulli, Perez-Cabello, & de la Riva, 2007):

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

where,

$n$  = number of observation points

$h$  = bandwidth

$K$  = central module (kernel)

$x$  = vector of coordinates representing the location where the function is estimated

$X$  = vectors of coordinates representing each observation point

$d$  = number of dimensions in space.

análisis de autocorrelación espacial (Anselin, Bera, Florax, & Yoon, 1996), b) análisis *clúster* (agrupación de puntos) (Kulldorff & Nagarwalla, 1995) y c) modelos de autorregresión espacial (Anselin et al., 1996); y 2) mapas de densidad, donde se usan básicamente, las técnicas de análisis de cuadrantes y la densidad kernel (Fuenzalida, Cobs, & Guerrero, 2013). Esta última se ha usado cuando se tienen datos históricos de incendios (Koutsias et al., 2004; Kuter et al., 2011). La densidad kernel es una técnica no paramétrica basada en varias funciones, algunas de estas son: cuadrática (Silverman, 1986), uniforme, Epanechnikov, distribución normal, función triangular y función cuártica (Turlach, 1999). De esta forma, el estimador de densidad kernel, para un caso multivariado, se define con base en la siguiente expresión matemática (Amatulli, Perez-Cabello, & de la Riva, 2007):

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

donde,

$n$  = número de puntos de observación

$h$  = ancho de banda

Based on the above, kernel density estimation helps to generate continuous surfaces of forest fire density. For that purpose, local neighborhood calculations are used under the structure of a grid (grid of cells), where the density value, at a given point or cell, results from an estimate based on the values of several points around it; nearby values have greater influence, while distant ones have less weight in the estimate. This conforms to the first law of geography, which states that everything is related to everything, but things that are close are more related to each other than things that are more distant; however, this process is limited to a certain distance, which is defined by establishing a bandwidth ( $h$ ) (de Smith, Goodchild, & Longley, 2009).

### Determining bandwidth ( $h$ )

Because the application of the kernel function requires that a bandwidth ( $h$ ) be specified, this study considers the processes described below.

1) Theoretical polygon. It is assumed a theoretical square surface, where the theoretical  $h$  is defined based on the length of the corresponding theoretical area (de la Riva et al., 2004):

$$r = D/2$$

where,

$D$  = diagonal of a theoretical square.

2) Mean random distance (MRD). The criteria used to define the MRD were expressed mathematically in the following way (de la Riva et al., 2004):

$$MRD = \frac{1}{2} \sqrt{\frac{A}{N}}$$

where,

$A$  = average size of the polygon ( $\text{km}^2$ )

$N$  = average number of ignition points within the polygon.

3) Average distance of the nearest neighbor. The  $h$  value is defined by estimating the average distance among nearest neighbor fires (Koutsias et al., 2014).

4) Silverman's rule of thumb. It considers a non-Gaussian kernel function, where the  $h$  value is defined as follows (Silverman, 1986):

$$h = \left( \frac{4\sigma^5}{3n} \right)^{1/5} \approx 1.06 \sigma n^{-1/5}$$

where,

$\sigma$  = standard deviation of sampled points

$n$  = number of sampled points.

$K$  = módulo central (kernel)

$x$  = vector de coordenadas que representa la ubicación donde la función es estimada

$X$  = vectores de coordenadas que representan cada punto de observación

$d$  = número de dimensiones en el espacio.

Con base en lo anterior, la estimación de densidad kernel ayuda a generar superficies continuas de la densidad de incendios forestales. Para ello, se usan cálculos de vecindad local que se hacen bajo la estructura de una cuadrícula (red de celdas), donde el valor de densidad, en un punto o celda dada, resulta de una estimación basada en los valores de varios puntos a su alrededor; los valores cercanos tienen mayor influencia, mientras que los lejanos tienen menor ponderación en la estimación. Esto se ajusta a la primera ley de la geografía, que señala que todo está relacionado con todo, pero las cosas que se encuentran cercanas están más relacionadas entre sí que las cosas que se encuentran más lejanas; sin embargo, este proceso se limita a cierta distancia, la cual se define al establecer un ancho de banda ( $h$ ) (de Smith, Goodchild, & Longley, 2009).

### Determinación de ancho de banda ( $h$ )

Debido a que la aplicación de la función kernel requiere que se especifique un ancho de banda ( $h$ ), para su definición, en este trabajo, se consideraron los procesos descritos a continuación.

1) Polígono teórico. Se asume una superficie cuadrada teórica, donde el  $h$  teórico se define con base en la longitud del radio teórico correspondiente (de la Riva et al., 2004):

$$r = D/2$$

donde,

$D$  = diagonal de un cuadrado teórico.

2) Distancia aleatoria media (DAM). El criterio usado para definir la DAM se expresó matemáticamente de la siguiente forma (de la Riva et al., 2004):

$$DAM = \frac{1}{2} \sqrt{\frac{A}{N}}$$

donde,

$A$  = tamaño medio del polígono ( $\text{km}^2$ )

$N$  = número medio de puntos de ignición dentro del polígono.

3) Distancia promedio del vecino más cercano. El valor de  $h$  se define con el cálculo de la distancia promedio que existe entre los incendios vecinos más cercanos (Koutsias et al., 2014).

In other cases, for example with a triangular kernel function, the  $h$  value would be estimated with the following equation:

$$h = 2.576 \sigma n^{-1/5}$$

5) Interquartile range. It assume that distribution deviates from normality and use the following algorithm to determine the value of  $h$  (Vrahimis, 2010):

$$h = 0.9 * \min \left( DS, \frac{\widehat{IQR}}{1.34} \right)$$

where,

$DS$  = standard distance

$IQR$  = interquartile range

### Validation criteria

The statistically most adequate  $h$  value was selected by the following criteria.

1) Root mean square error (RMSE). This statistic was determined in terms of the difference between estimated density and true density, where the lowest RMSE was selected. First, we calculated the mean squared error (MSE) that incorporates both the variance of the estimator and its bias and, later, the square root of MSE was obtained:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\widehat{y}_1 - y_1)^2}$$

where,

$n$  = number of points compared

$\widehat{y}_1$  = estimated density

$y_1$  = true density.

Based on this, 1 000 random points were located in the study area to determine RMSE; and with them two common surfaces (circular area) were compared: 10 and 100 km<sup>2</sup>. The real value of the number of fires from 2005-2013 and the estimated value of the number of fires defined by the kernel function were determined in these areas.

2) Root mean integrated squared error. The comparison was made under the same criteria indicated for the RMSE. For this, the mean integrated squared error (MISE) was estimated and then the square root (RMISE) was calculated, from the equation proposed by Seaman and Powell (1996):

$$RMISE = \sqrt{\frac{1}{n} \sum_{i=1}^n \frac{(\widehat{y}_1 - y_1)^2}{y_1}}$$

4) Regla de oro de Silverman ("Silverman's rule of thumb"). Considera una función kernel no gaussiana, donde el valor  $h$  se define de la siguiente forma (Silverman, 1986):

$$h = \left( \frac{4\sigma^5}{3n} \right)^{1/5} \approx 1.06 \sigma n^{-1/5}$$

donde,

$\sigma$  = desviación estándar de los puntos muestreados

$n$  = número de puntos muestreados.

En otros casos, por ejemplo, con una función kernel triangular, el valor de  $h$  se estimaría con la siguiente ecuación:

$$h = 2.576 \sigma n^{-1/5}$$

5) Rango intercuartílico. Asume que la distribución se desvía de la normalidad y usa el siguiente algoritmo para determinar el valor de  $h$  (Vrahimis, 2010):

$$h = 0.9 * \min \left( DS, \frac{\widehat{IQR}}{1.34} \right)$$

donde,

$DS$  = distancia estándar

$IQR$  = rango intercuartílico

### Criterios de validación

El valor de  $h$  estadísticamente más adecuado se seleccionó mediante los siguientes criterios.

1) Raíz del error cuadrático medio (RECM). Este estadístico se determinó en términos de la diferencia entre la densidad estimada y la densidad real, donde se seleccionó el menor RECM. Primero se calculó el error cuadrático medio (ECM) que incorpora tanto la varianza del estimador como su sesgo y, posteriormente, se obtuvo la raíz cuadrada del ECM:

$$RECM = \sqrt{\frac{1}{n} \sum_{i=1}^n (\widehat{y}_1 - y_1)^2}$$

donde,

$n$  = número de puntos comparados

$\widehat{y}_1$  = densidad estimada

$y_1$  = densidad real.

Con base en esto, para determinar el RECM se ubicaron 1 000 puntos al azar en el área de estudio, con los que se compararon dos superficies comunes (área circular): 10 y 100 km<sup>2</sup>. En estas áreas se determinaron el valor real del número de incendios del periodo 2005-2013 y el valor estimado del número de incendios definido por la función kernel.

where,

$n$  = number of points compared

$\hat{y}_1$  = estimated density

$y_1$  = true density.

3) Coefficient of variation (CV). Taking into account that the size of  $h$  defines a sampling unit allowing to capture the variability of the number of forest fires, several circular surfaces or reference sites (CRS) were analyzed in the study area: 1, 2, 4, 8, 10, 15, 30, 50, 70, 100, 150 and 200 km<sup>2</sup>. In addition, the following sampling intensities completely randomly distributed were evaluated: 100, 300, 500 and 1000 SIR. The contained fires in each SIR were counted and, then, the CV defined in relation to the 12 sizes of SIR was plotted; in this way, the surface of SIR and its corresponding  $h$  were selected at the point where the variability trend starts an asymptotic behavior.

4) Comparative percentage. Similar to the CV, the increase in  $h$  eventually defines an asymptotic behavior of RMISE, so, an objective selection of  $h$  is to locate the break point or at least a break rank, where this behavior starts. For this, the percentage representing the value of RMISE was estimated in relation to the immediate superior value, which was called comparative percentage (CP), which is calculated with the following equation:

$$CP = \left( \frac{RMISE_n}{RMISE_{n-1}} \right) * 100$$

where,

$RMISE_n$  = root mean integrated squared error at a given  $h$  value (bandwidth)

$RMISE_{n+1}$  = root mean integrated squared error of the  $h$  value immediate to that determined in  $RMISE_n$ .

## Results and discussion

### Estimating $h$

Table 1 shows the bandwidth estimates ( $h$  = search area), resulting from the processes analyzed. There is great variation in the  $h$  values obtained; the range between the maximum and the minimum value was 39 356.34 m with an average of  $10\ 936.74 \pm 9\ 955.04$  m. Table 1 also shows the values of the criteria on which the analyzed processes are based. It is important to note that some of the criteria are particular to certain processes, so they were not used in all cases.

In the process of the theoretical polygon, the selection of  $h$  is according to the scale on which the information is held; that is, if the information is presented in pixels of  $10 \times 10$  km, then a theoretical area is established corresponding to half the length of the frame defining

2) Raíz del error medio cuadrático integrado. La comparación se hizo bajo los mismos criterios señalados para la RECM. Para ello, primeramente, se determinó el error medio cuadrático integrado (EMCI) y después se calculó su raíz cuadrada (REMCI), a partir de la ecuación propuesta por Seaman y Powell (1996):

$$REMCI = \sqrt{\frac{1}{n} \sum_{i=1}^n \frac{(\hat{y}_1 - y_1)^2}{y_1}}$$

donde,

$n$  = número de puntos comparados

$\hat{y}_1$  = densidad estimada

$y_1$  = densidad real.

3) Coeficiente de variación (CV). Considerando que el tamaño de  $h$  define una unidad de muestreo que permite captar la variabilidad del número de incendios forestales, en el área de estudio se analizaron varias superficies circulares o sitios de referencia (SIR): 1, 2, 4, 8, 10, 15, 30, 50, 70, 100, 150 y 200 km<sup>2</sup>. Además, se evaluaron las siguientes intensidades de muestreo distribuidas completamente al azar: 100, 300, 500 y 1000 SIR. Los incendios contenidos se contabilizaron en cada SIR y, posteriormente, se graficó el CV que se define con relación a los 12 tamaños de SIR; de esta forma, la superficie del SIR y su  $h$  correspondiente se seleccionaron en el punto donde la tendencia de la variabilidad inicia un comportamiento asintótico.

4) Porcentaje comparativo. En forma similar al CV, el aumento de  $h$  eventualmente define un comportamiento asintótico de REMCI, por lo que, de igual forma, una selección objetiva de  $h$  es ubicar el punto de quiebre o al menos un rango de quiebre, donde se inicia dicho comportamiento. Para esto, el porcentaje que representa el valor de REMCI se determinó con relación al valor superior inmediato, el cual se denominó porcentaje comparativo (PC), mismo que se calcula con la siguiente ecuación:

$$PC = \left( \frac{REMCI_n}{REMCI_{n-1}} \right) * 100$$

donde,

$REMCI_n$  = raíz de error medio cuadrático integrado en un valor de  $h$  (ancho de banda) dado

$REMCI_{n-1}$  = raíz de error medio cuadrático integrado del valor de  $h$  inmediato al determinado en  $REMCI_n$ .

## Resultados y discusión

### Estimación de $h$

El Cuadro 1 presenta las estimaciones del ancho de banda (radio de búsqueda  $h$ ), resultantes de los

the pixels (de la Riva et al., 2004). However, although this alternative may be practical, it is not based on statistics; for example, it should be considered that the variance of the number of fires can change if one works with different areas defined by multiples or submultiples of pixel size. As an alternative, it was proposed to base the definition of  $h$  on the average size of polygons, assuming a polygon of 20 km<sup>2</sup> as the theoretical square surface, where the theoretical  $h$  (2 550 m) was defined based on the length of the corresponding theoretical area (de la Riva et al., 2004), which resulted in the lowest  $h$  value of all the processes analyzed.

On the other hand, the average distance defined among 1 000 points located randomly (MRD), was used as  $h$  value, which can be shown from two perspectives (Kuter et al., 20011): a) local, regarding the average polygon

procesos probados. Existe gran variación en los valores  $h$  obtenidos; el rango entre el valor máximo y el mínimo fue 39 356.34 m con una media de 10 936.74  $\pm$  9 955.04 m. En el Cuadro 1 también se muestran los valores de los criterios en los que se basan los procesos probados. Es importante señalar que algunos de los criterios son particulares de ciertos procesos, por lo que no se usaron en todos.

En el proceso del polígono teórico, la selección de  $h$  está en función de la escala en la que se tenga la información; es decir, si la información se presenta en píxeles de 10 x 10 km, entonces se establece un radio teórico que corresponde a la mitad de la longitud del cuadro que define los píxeles (de la Riva et al., 2004). No obstante, aunque esta alternativa puede ser práctica, no se fundamenta de manera estadística; por ejemplo, debe

**Table 1. Bandwidth ( $h$ ) estimation corresponding to the kernel density estimator, according to different processes. Cuadro 1. Estimación del ancho de banda ( $h$ ), correspondiente al estimador de densidad kernel, de acuerdo con diferentes procesos.**

| Process / Proceso   | Source / Fuente          | H (m)  | P (km <sup>2</sup> ) | A (km <sup>2</sup> ) | N    | Particular criteria / Criterios particulares                  |
|---|--------------------------|--------|----------------------|----------------------|------|---|
| Theoretical polygon / Polígono teórico  | de la Riva et al. (2004) | 2 550  | 20                   | 13                   |      | Side / Lado = 3 605.55 m<br>D = 5 099.02 m                    |
| Mean random distance / Distancia aleatoria media                              | de la Riva et al. (2004) | 3 472  | 38                   | 5                    | 0.52 | A = 25, 100, 225 and 400 km <sup>2</sup>                      |
|   |                          | 5 395  | 91                   | 10                   | 0.86 |   |
|   |                          | 6 750  | 143                  | 15                   | 1.23 |   |
|   |                          | 7 797  | 191                  | 20                   | 1.65 |   |
| Twice the mean random distance / Doble de la distancia aleatoria media        | Kuter et al. (2011)      | 6 944  | 151                  | 5                    | 0.52 | A = 25, 100, 225 and 400 km <sup>2</sup>                      |
|   |                          | 10 790 | 366                  | 10                   | 0.86 |   |
|   |                          | 13 500 | 573                  | 15                   | 1.23 |   |
|   |                          | 15 594 | 764                  | 20                   | 1.65 |   |
| Average distance to nearest fire / Distancia promedio al incendio más cercano | Koutsias et al. (2004)   | 4 980  | 78                   |                      |      |   |
| Silverman's Golden Rule / Regla de oro de Silverman                           | Silverman (1986)         | 9 051  | 257                  |                      |      | Factor = 1.06<br>DS = 46 562<br>n = 4 821                     |
| Triangular kernel function / Función kernel triangular                        | Turlach (1999)           | 21 996 | 1 520                |                      |      | Factor = 2.576<br>DS = 46 562<br>n = 4 821                    |
| Interquartile range / Rango intercuartílico                                   | Vrahimis (2010)          | 41 906 | 5 517                |                      |      | IQR = 64 824<br>IQR/1.34 = 48 376<br>DS = 46 562<br>n = 4 821 |

$h$  = search area, DS = standard deviation of the sampled points, A = average polygon size, D = diagonal (m), N = average number of forest fires, n = number of sampled points, P = size of the sampling polygon (km<sup>2</sup>), IQR = interquartile range.

$h$  = radio de búsqueda, DS = desviación estándar de los puntos muestreados, A = tamaño promedio de polígono, D = diagonal (m), N = número promedio de incendios forestales, n = número de puntos muestreados, P = tamaño del polígono de muestreo (km<sup>2</sup>), IQR = rango intercuartílico.

size and a number of ignition points per polygon, and b) global, where the total size of the study area and the total number of ignition points are considered. In this case, the first perspective was considered, since it allowed to establish the value of  $h$  in relation to the different sizes of polygons that were analyzed (25, 100, 225 and 400 km<sup>2</sup>); however, the differences among the corresponding  $h$  values were not so considerable, with values of 3 472, 5 395, 6 750 and 7 797 m, respectively. Therefore, based on that recommended by de la Riva et al. (2004), twice the mean random distance was also analyzed as the  $h$  value; however, neither the MRD nor its double version work directly with the points defining the location of forest fires.

To understand the problematic of the selection of  $h$ , it must be understood that it is a variable (within a mathematical model) that can be defined in fixed or adaptive mode. Regarding the first option,  $h$  was also defined based on the concept of neighbor's average distance, where the nearest fire point was taken (Koutsias et al., 2004). In this way, the resulting value was lower than the average of MRD defined (5 854).

One of the first approaches to a selection criteria of the  $h$  value is based on the "Golden rule of Silverman", which assumes a kernel function with a Gaussian base, to approximate univariate data, where it is suggested to select the  $h$  value minimizing the mean integrated squared error (risk function) (Turlach, 1999). In this study, the  $h$  value resulting from this perspective was 9 051 m, which approaches to the criteria average of twice the MRD analyzed (11 707); however, regarding a triangular kernel function, the value of the resulting  $h$  was more than double (21 996).

On the other hand, considering that the distribution deviates from normality, the implementation of the interquartile range algorithm (Silverman, 1986) resulted in an extremely overestimated  $h$  value. This coincides with that found by Silverman (1986) in a case of a bimodal distribution with separate means, defining a highly skewed distribution.

### Continuous surfaces

Once the  $h$  values were defined, the kernel function estimator was implemented, defining two options for the generation of continuous surfaces: a) based on the estimated density probabilities; and b) based on the estimate of the corresponding density (number of fires per unit area). Taking into account that in the definition of fire management strategies it is necessary to have mapping to locate and size priority areas, the continuous surfaces were generated based on the density estimated with the different  $h$  values. Figure 2

considerarse que la varianza del número de incendios puede cambiar si se trabaja con áreas diferentes que se definen por múltiplos o submúltiplos del tamaño del pixel. Como alternativa, se propuso basar la definición de  $h$  en el tamaño promedio de polígonos, asumiendo un polígono de 20 km<sup>2</sup> como la superficie cuadrada teórica, donde el  $h$  teórico (2 550 m) se definió con base en la longitud del radio teórico correspondiente (de la Riva et al., 2004), lo cual resultó en el menor valor de  $h$  de todos los procesos probados.

Por otra parte, el promedio de la distancia, que se definió entre 1 000 puntos ubicados al azar (DAM), se usó como valor de  $h$ , el cual puede presentarse desde dos perspectivas (Kuter et al., 20011): a) local, considerando el tamaño medio de polígono y un número de puntos de ignición por polígono, y b) global, donde se considera el tamaño total del área de estudio y el número total de puntos de ignición. En este caso, se consideró la primera perspectiva, ya que permitió establecer el valor de  $h$  con relación a los diferentes tamaños de polígonos que se probaron (25, 100, 225 y 400 km<sup>2</sup>); sin embargo, las diferencias entre los valores de  $h$  correspondientes no fueron tan marcadas, con valores de 3 472, 5 395, 6 750 y 7 797 m, respectivamente. Por tanto, con base en lo recomendado por de la Riva et al. (2004), el doble de la distancia aleatoria media se probó también como el valor de  $h$ ; sin embargo, ni la DAM ni su versión duplicada trabajan directamente con los puntos que definen la ubicación de los incendios forestales.

Para comprender la problemática que significa la selección de  $h$ , debe entenderse que es una variable (dentro de un modelo matemático) que puede definirse en modo fijo o en modo adaptativo. Considerando la primera opción, también se definió  $h$  con base en el concepto de la distancia promedio del vecino, donde se tomó el punto de incendio más cercano (Koutsias et al., 2004). De esta manera, el valor resultante fue menor que el promedio de las DAM definidas (5 854).

Una de las primeras aproximaciones a un criterio de selección del valor de  $h$  es con base en la "regla de oro de Silverman", la cual asume una función kernel de base gaussiana, para aproximar datos univariados, donde se sugiere seleccionar el valor de  $h$  que minimice el error medio cuadrático integrado (función de riesgo) (Turlach, 1999). En este trabajo, el valor de  $h$  resultante bajo esta perspectiva fue de 9 051 m, el cual se aproxima al promedio del criterio del doble de las DAM probadas (11 707); sin embargo, considerando una función kernel triangular, el valor de  $h$  resultante fue más del doble (21 996).

Por otra parte, considerando que la distribución se desvía de la normalidad, la implementación del



shows the continuous surfaces, observing that the surface “softens” as the value of  $h$  increases, while, at low  $h$  values, the continuous surface tends to define circular regions around the points (forest fires), what is known as the “bull’s eye” phenomenon.

algoritmo de rango intercuartílico (Silverman, 1986) resultó en un valor de  $h$  extremadamente sobrestimado. Esto coincide con lo encontrado por Silverman (1986) en un caso de una distribución bimodal con medias separadas, que define una distribución altamente sesgada.

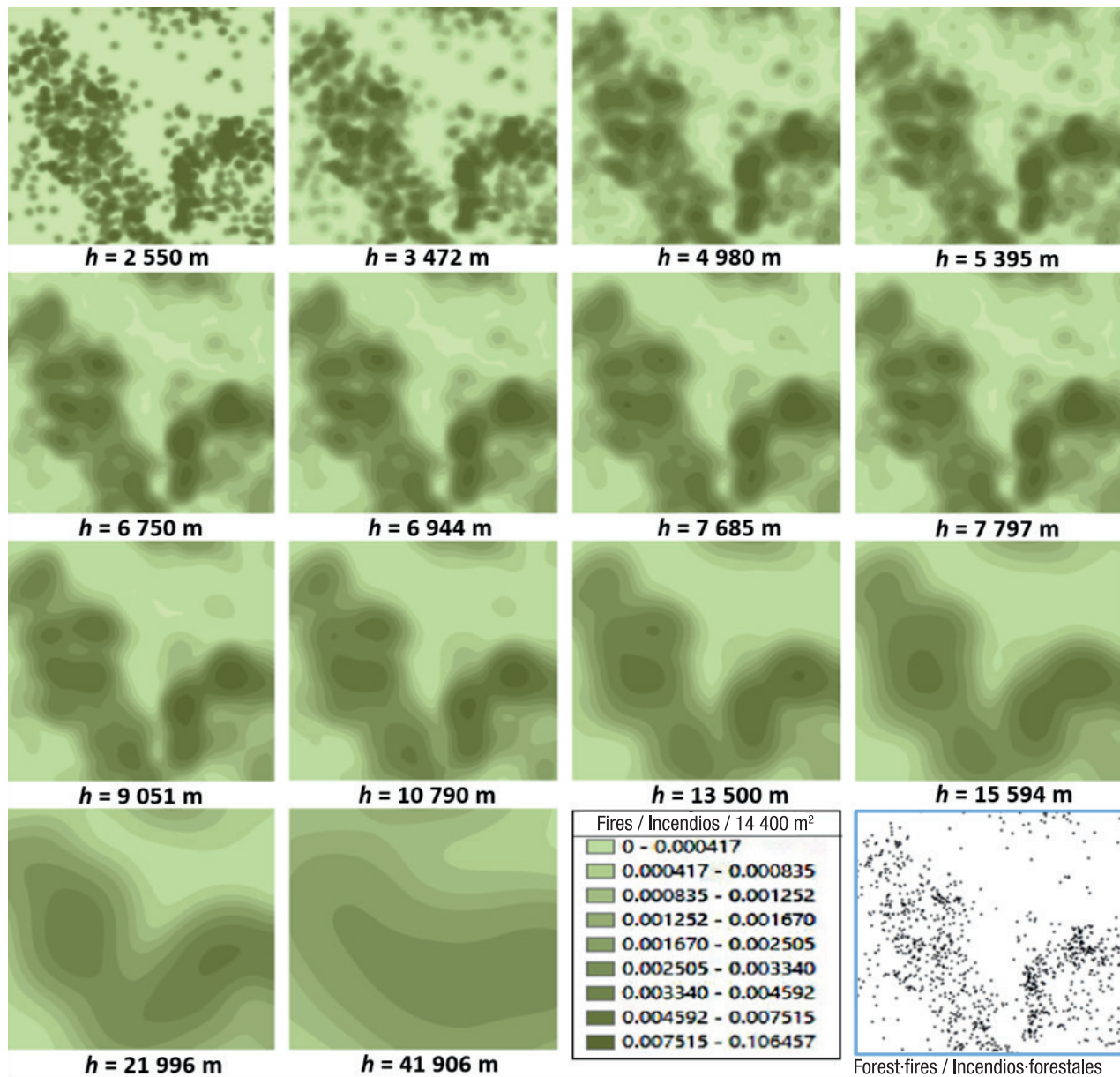


Figure 2. Continuous surfaces of forest fire density (pixels of 14 400 m<sup>2</sup> [120 x 120 m]), derived from the kernel density function with different bandwidth values ( $h$ ).

Figura 2. Superficies continuas de la densidad de incendios forestales (píxeles de 14 400 m<sup>2</sup> [120 x 120 m]), derivadas de la función de densidad kernel con diferentes valores de ancho de banda ( $h$ ).

### Selecting bandwidth ( $h$ )

An option for selecting the  $h$  value is to do it intuitively, based on a visual perspective (de la Riva et al., 2004; Koutsias et al., 2004; Krisp et al., 2009); that is, subjectively selecting the  $h$  value producing a continuous surface, whose softness is more “random” than “structural” (Kuter et al., 2011). Therefore, it is considered that this selection is an act of art rather than science (Krisp et al., 2009), which does not allow to define replicable processes and will always be subject to the experience of the person making such selection (Turlach, 1999). As an alternative, this study implemented statistical criteria, seeing in Figure 3 that when comparing continuous surfaces (generated by the different  $h$  values), errors tend to increase as the  $h$  value increases. This trend is adjusted to a second-order polynomial regression (Table 2), without defining an asymptotic behavior, so it is not possible to establish a breakpoint that allows defining a maximum  $h$  value (or a range). This suggested that a common comparison area should be established to analyze forest fire density estimations.

According to the above, regarding the RMSE and a common comparison surface of 10 km<sup>2</sup>, Figure 4 shows that when the  $h$  value increases, the estimation error also increases, and as it was observed in Figure 2, the “smoothing” of the continuous surfaces increases in proportion to the increase in the  $h$  value. This implies that the selection of a smoothed continuous surface (de la Riva et al., 2004; Koutsias et al., 2004; Kuter et al., 2011) does not necessarily define the best  $h$  value. Table 2 shows the statistics corresponding to this trend. Whereas the utility of the RMSE formula can be

### Superficies continuas

Una vez definidos los valores de  $h$ , se procedió a implementar el estimador de la función kernel, con el cual se definen dos opciones para la generación de superficies continuas: a) con base en las probabilidades de densidad estimada; y b) con base en la estimación de la densidad (número de incendios por unidad de superficie) correspondiente. Considerando que en la definición de estrategias de manejo del fuego se requiere contar con cartografía que permita ubicar y dimensionar áreas prioritarias, las superficies continuas se generaron con base en la densidad estimada con los diferentes valores de  $h$ . La Figura 2 muestra las superficies continuas, observándose que la superficie se “suaviza” a medida que el valor de  $h$  incrementa, mientras que, en valores bajos de  $h$ , la superficie continua tiende a definir regiones circulares alrededor de los puntos (incendios forestales), lo que se conoce como el fenómeno de “ojo de buey”.

### Selección de ancho de banda ( $h$ )

Una opción para la selección del valor de  $h$  es hacerla en forma intuitiva, con base en una perspectiva visual (de la Riva et al., 2004; Koutsias et al., 2004; Krisp et al., 2009); es decir, subjetivamente se selecciona el valor de  $h$  que produzca una superficie continua, cuya suavidad sea más “aleatoria” que “estructural” (Kuter et al., 2011). Por lo anterior se considera que esta selección es un acto más de arte que de ciencia (Krisp et al., 2009), lo cual no permite definir procesos replicables y estará siempre supeditado a la experiencia de la persona que haga tal selección (Turlach, 1999). Como alternativa, en este trabajo se implementaron criterios

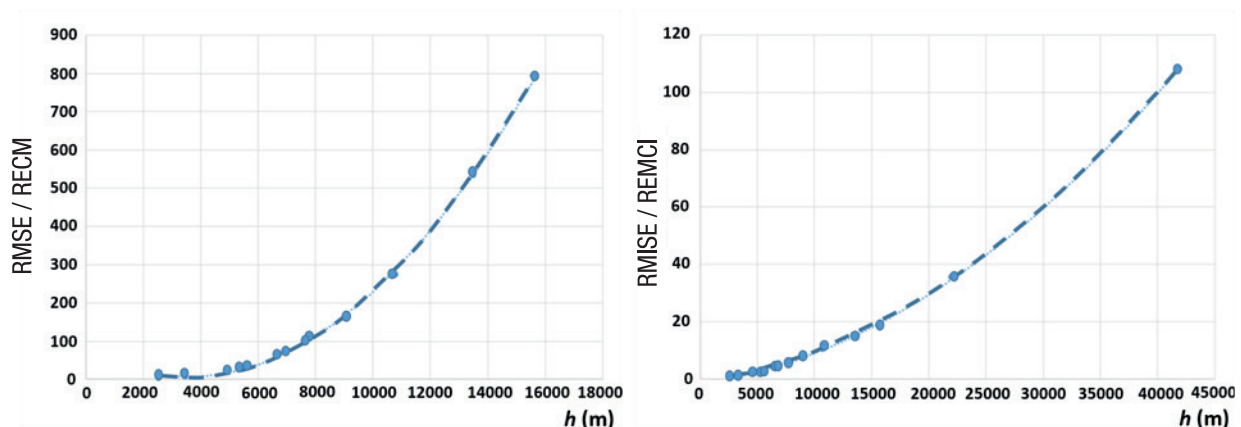


Figure 3. Error distribution in relation to bandwidth ( $h$ ), when comparing forest fire density estimations in areas defined by different  $h$ . RMSE = root mean square error, RMISE = root mean integrated squared error.

Figura 3. Distribución de errores en relación con el ancho de banda ( $h$ ), al comparar estimaciones de densidad de incendios en áreas definidas por diferentes  $h$ . RECM = raíz de error cuadrático medio, REMCI = raíz de error medio cuadrático integrado.

**Table 2. Trend models between the relationship of statistical selection criteria and bandwidth value ( $h$ ).****Cuadro 2. Modelos de la tendencia entre la relación de los criterios estadísticos de selección y el valor del ancho de banda ( $h$ ).**

| Comparative / Comparativo | Criteria / Criterio | Model / Modelo                       | $r^2$  |
|---------------------------|---------------------|--------------------------------------|--------|
| 10 km <sup>2</sup>        | RMSE / RECM         | = 0.3638ln( $h$ ) - 2.2248           | 0.9452 |
|                           | RMISE / REMCI       | = 0.0465ln( $h$ ) - 0.1052           | 0.7823 |
| 100 km <sup>2</sup>       | RMSE / RECM         | = 0.7927ln( $h$ ) + 5.1792           | 0.8135 |
|                           | RMISE / REMCI       | = 0.1977ln( $h$ ) + 0.4507           | 0.4437 |
| $h$                       | RMSE / RECM         | = 5E-06 $h^2$ - 0.0381 $h$ + 75.0210 | 0.9992 |
|                           | RMISE / REMCI       | = 5E-08 $h^2$ + 0.0006 $h$ - 1.7225  | 0.9999 |

RMSE = root mean square error; RMISE= root mean integrated squared error.

RECM = raíz cuadrada del error cuadrático medio; REMCI = raíz cuadrada del error medio cuadrático integrado.

doubtful in practice, since it depends on the  $f$  unknown density, the RMISE was used as a global measure of the estimator's precision (Álvarez & Yohai, 2012). The trend between RMISE and  $h$  was similar; however, the adjustment model had a lower  $r^2$  value (Table 2), showing an asymptotic behavior when  $h = 15\ 000$  m.

In the case of the common area of 100 km<sup>2</sup>, the errors of RMSE were greater and very similar among the different  $h$  values. This trend suggests that, as the area of comparison increases, the error of the estimates tends to be more homogeneous. Regarding the RMISE, the behavior was similar to that of RMSE; nevertheless, it is observed that, in low  $h$  values, there is a marked tendency to reduce the error. On the other hand, in both cases (RMSE and RMISE) an asymptotic tendency is defined from  $h$  values lower than 5 000 m. The comparison of RMISE, regarding the areas of 10 and 100 km<sup>2</sup>, suggests that the  $h$  value, with constant behavior of the error, is between 5 000 and 10 000 m, where the values defined in the processes are shown in Table 1: mean random distance ( $h = 5\ 395, 6\ 750, 7\ 797$  m), twice the mean random distance ( $h = 6\ 944$  m), average distance to the nearest fire and "Silverman's golden rule" (Silverman, 1986).

According to the results, the size of  $h$  with the lowest estimation error (RMISE) should be chosen, which corresponds to the lowest  $h$  value. However, this does not guarantee that we have the most adequate  $h$  value, since it should also be considered that the variation (CV) of forest fire density decreases as the size of  $h$  increases (Figure 5). However, the choice of low CV values does not guarantee the choice of the most suitable  $h$  value.

estadísticos, observando en la Figura 3 que al comparar las superficies continuas (generadas por los diferentes valores de  $h$ ), los errores tienden a incrementarse a medida que el valor de  $h$  aumenta. Esta tendencia se ajusta a una regresión polinómica de segundo orden (Cuadro 2), sin llegar a definir un comportamiento asintótico, por lo que no es posible establecer un punto de quiebre que permita definir un valor (o un rango) de  $h$  máximo. Esto sugirió que se debe establecer una superficie de comparación común para analizar las estimaciones de la densidad de incendios forestales.

De acuerdo con lo anterior, considerando la RECM y una superficie de comparación común de 10 km<sup>2</sup>, en la Figura 4 se observa que a medida que el valor de  $h$  aumenta, el error de estimación también lo hace, y como se observó en la Figura 2, la "suavización" de las superficies continuas aumenta en proporción con el incremento del valor de  $h$ . Esto implica que la selección de una superficie continua suavizada (de la Riva et al., 2004; Koutsias et al., 2004; Kuter et al., 2011) no necesariamente define el mejor valor de  $h$ . El Cuadro 2 presenta los estadísticos que corresponden a esta tendencia. Considerando que la utilidad de la fórmula del RECM puede ser dudosa en la práctica, ya que depende de la densidad desconocida  $f$ , el REMCI se usó como una medida global de la precisión del estimador (Álvarez & Yohai, 2012). La tendencia entre REMCI y  $h$  fue similar; sin embargo, el modelo de ajuste presentó un valor menor de  $r^2$  (Cuadro 2), mostrando un comportamiento asintótico aproximadamente cuando  $h = 15\ 000$  m.

En el caso de la superficie común de 100 km<sup>2</sup>, los errores de la RECM fueron mayores y muy parecidos entre los

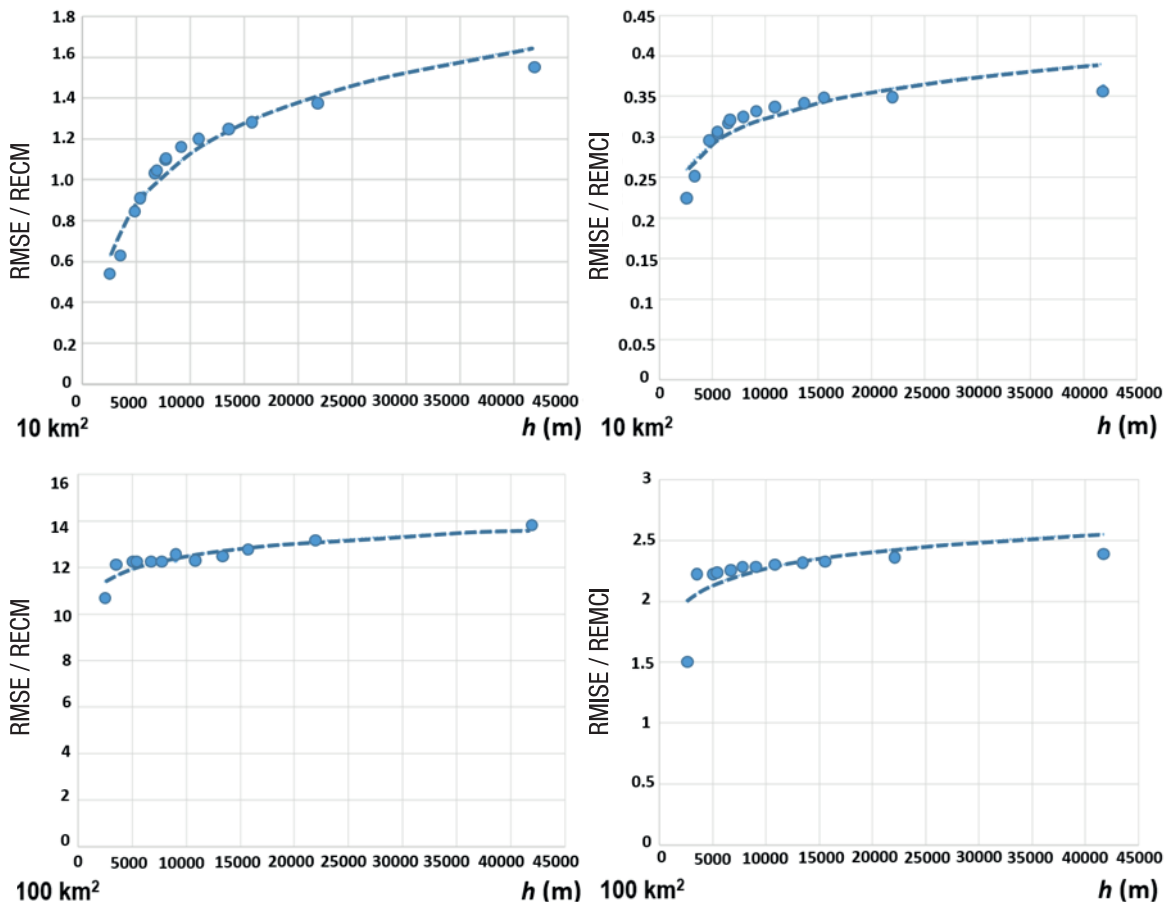


Figure 4. Estimation error distribution of forest fire density in relation to bandwidth ( $h$ ), regarding two comparison areas. RMSE = root mean square error; RMISE = root mean integrated squared error.

Figura 4. Distribución de errores de estimación de la densidad de incendios forestales en relación con el ancho de banda ( $h$ ), considerando dos áreas de comparación. RECM = raíz de error cuadrático medio; REMCI = raíz de error medio cuadrático integrado.

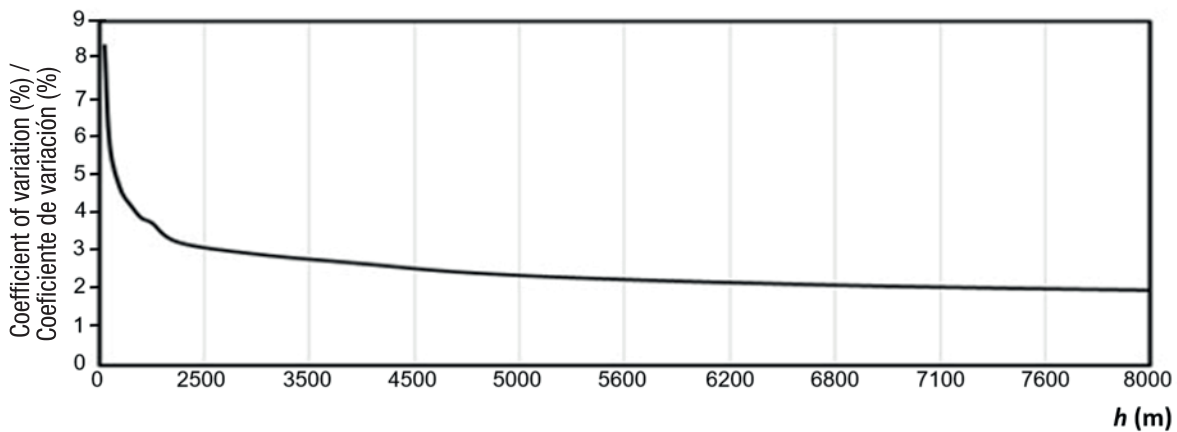


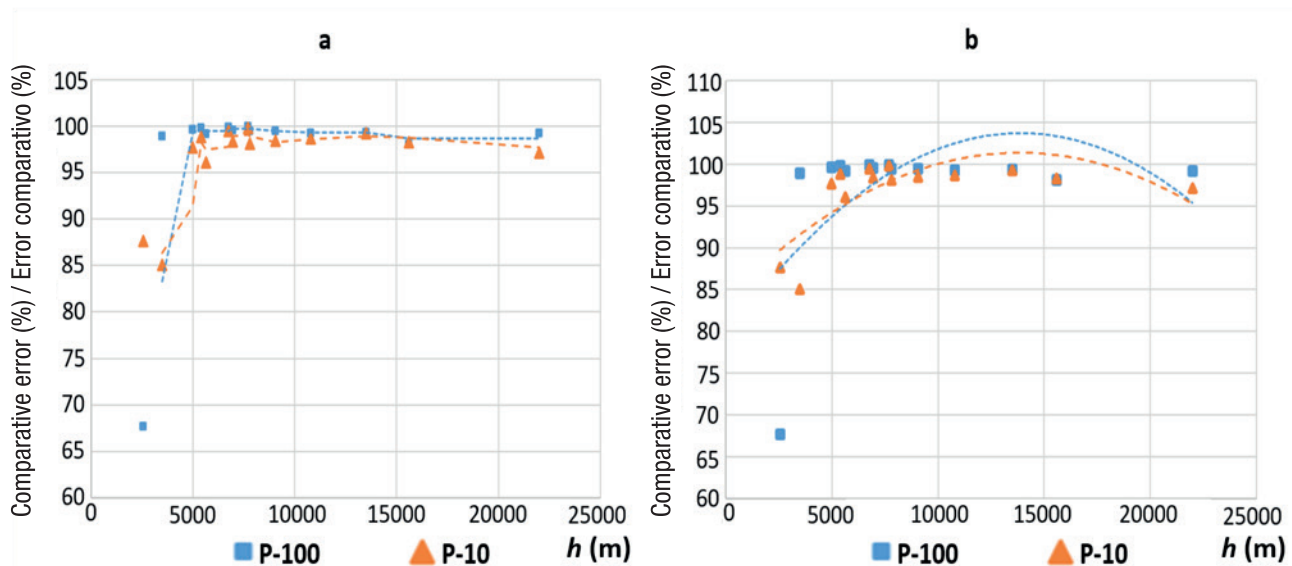
Figure 5. Coefficient of variation trend of forest fire density in relation to the increase in bandwidth ( $h$ ).

Figura 5. Tendencia del coeficiente de variación de la densidad de incendios forestales en relación con el aumento del ancho de banda ( $h$ ).

Based on the above, it was found that variance (CV) decreases when  $h$  value increases, while the bias (RMISE) increases; the opposite occurs when  $h$  value decreases. Therefore, the definition of the most adequate  $h$  value will be that which represents a compromise between variance and bias (Álvarez & Yohai, 2012). For this, based on the asymptotic behavior of RMISE, an alternative way is defined to determine the  $h$  value, by locating the break point or at least a break rank where the asymptote starts. For this purpose, the comparative percentage (CP) values were determined, where a break range is defined, around 5 000 and 6 000 m (Figure 6a), this situation coincides for the two polygon sizes of comparison (10 and 100 km<sup>2</sup>). In order to locate a breaking point, the CP trends were modeled (P-10:  $y = -9E-08x^2 + 0.0025x + 83.87$ ,  $[R^2 = 0.5891]$ ; P-100:  $y = -1E-07x^2 + 0.0035x + 79.358$ ,  $[R^2 = 0.5891]$ ), whose crossing defined a point of coincidence around 5 600 m ( $h$ ), which is within the breaking range defined (Figure 6b). Finally, the statistically most adequate  $h$  value will be the one that comes closest to 5 600 m and which is within the range of 5 000 and 6 000 m. According to this, the definition process of  $h$  that is closest to this value is the mean random distance ( $h = 5 395$  m).

diferentes valores de  $h$ . Esta tendencia sugiere que, a medida que el área de comparación aumenta, el error de las estimaciones tiende a ser más homogéneo. Con relación al REMCI, el comportamiento fue similar al del RECM; no obstante, se observa que, en valores bajos de  $h$ , hay una tendencia marcada a reducir el error. Por otra parte, en ambos casos (RECM y REMCI) se llega a definir una tendencia asintótica a partir de valores de  $h$  menores de 5 000 m. La comparación de REMCI, considerando las áreas de 10 y 100 km<sup>2</sup>, sugiere que el valor de  $h$  en el cual se tiende a un comportamiento constante del error se encuentra entre 5 000 y 10 000 m, donde se incluyen los valores definidos en los procesos presentados en el Cuadro 1: distancia aleatoria media ( $h = 5 395, 6 750, 7 797$  m), doble de la distancia aleatoria media ( $h = 6 944$  m), distancia promedio al incendio más cercano y “regla de oro de Silverman” (Silverman, 1986).

De acuerdo con los resultados, se debería elegir el tamaño de  $h$  con el menor error de estimación (REMCI), lo cual corresponde al valor más bajo de  $h$ . Sin embargo, esto no garantiza que se tenga el valor de  $h$  más adecuado, ya que también se debe considerar que la variación (CV) de la densidad de incendios forestales disminuye a medida que el tamaño de  $h$



**Figure 6. Comparative error in reference to immediate superior RMISE (root mean integrated squared error) and variations of  $h$  (bandwidth): a) moving average and b) polynomial model (second order). P-10 and P-100 = comparison polygons of 10 and 100 km<sup>2</sup>, respectively.**

**Figura 6. Error comparativo en referencia al REMCI (raíz de error medio cuadrático integrado) superior próximo y variaciones de  $h$  (ancho de banda): a) media móvil y b) modelo polinómico (segundo orden). P-10 y P-100 = polígonos de comparación de 10 y 100 km<sup>2</sup>, respectivamente.**

## Conclusions

It is possible to establish a statistical process of bandwidth ( $h$ ) selection, for kernel density estimation of forest fires. In this way, we tend to define processes and results, which can be replicable, comparable and compatible, which are not conditioned to the participation of experts, which avoids the use of subjective criteria such as the visual appreciation of continuous surfaces, defining the spatial variation of forest fire density. There was great variation among bandwidth values, which implies that there is no single and universal process for all cases. This is because kernel density estimation are conditioned to several factors, such as spatial resolution of information (cell size), bandwidth, number of points to be considered and intrinsic characteristics of the phenomenon under study.

### End of English version

## References / Referencias

- Álvarez, E., & Yohai, V. (2012). M-estimators for isotonic regression. *Journal of Statistical Planning and Inference*, 142(8), 2351–2368. doi: 10.1016/j.jspi.2012.02.051
- Amatulli, G., Perez-Cabello, F., & de la Riva, J. (2007). Mapping lightning/human caused wildfires occurrence under ignition point location uncertainty. *Ecological Modelling*, 200(3-4), 321–333. doi: 10.1016/j.ecolmodel.2006.08.001
- Anselin, L., Bera, A. K., Florax, R., & Yoon, M. J. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, 26(1), 77–104. doi: 10.1016/0166-0462(95)02111-6
- Comisión Nacional Forestal (CONAFOR). (2015). Incendios. Retrieved August 4, 2015 from <http://www.conafor.gob.mx/web/temas-forestales/incendios/>
- de la Riva, J., Pérez, C. F., Lana, R. N., & Koutsias, N. (2004). Mapping wildfire occurrence at regional scale. *Remote Sensing of Environment*, 92(3), 363–369. doi: 10.1016/j.rse.2004.06.022.
- de Smith, M. J., Goodchild, M. F., & Longley, P. A. (2009). *Geospatial analysis. A comprehensive guide to principles, Techniques and software tools* (3rd ed.). Winchelsea, UK: The Winchelsea Press. Retrieved from <http://www.spatialanalysisonline.com/>
- Fuenzalida, M., Cobs, V., & Guerrero, R. (2013). El territorio como unidad de análisis en estudios sobre las determinantes sociales de la salud. *Argos*, 30(59), 87–106. Retrieved from [http://www.academia.edu/27985206/El\\_territorio\\_como\\_unidad\\_de\\_an%C3%A1lisis\\_en\\_estudios\\_sobre\\_las\\_determinantes\\_sociales\\_de\\_la\\_salud](http://www.academia.edu/27985206/El_territorio_como_unidad_de_an%C3%A1lisis_en_estudios_sobre_las_determinantes_sociales_de_la_salud)
- Koutsias, N., Kalabokidis, K., & Allgöwer, B. (2004). Fire occurrence patterns at landscape level: Beyond positional accuracy of ignition points with kernel density estimation methods. *Natural Resource Modeling*, 17(4), 359–375. doi: 10.1111/j.1939-7445.2004.tb00141.x
- umenta (Figura 5). No obstante, tampoco la elección de valores de CV bajos garantiza la elección de valor de  $h$  más adecuado.
- Con base en lo anterior, se encontró que la varianza (CV) disminuye al aumentar el valor de  $h$ , mientras que el sesgo (REMCI) aumenta; ocurre lo contrario al disminuir el valor de  $h$ . Por tanto, la definición del valor más adecuado de  $h$  será aquel que represente un compromiso entre varianza y sesgo (Álvarez & Yohai, 2012). Para ello, con base en el comportamiento asintótico de REMCI, se define una alternativa para determinar el valor de  $h$ , al ubicar el punto de quiebre o al menos un rango de quiebre donde se inicia la asíntota. Para esto, se determinaron los valores de porcentaje comparativo (PC), con base en el cual se define un rango de quiebre, alrededor de los 5 000 y 6 000 m (Figura 6a), situación que coincide para los dos tamaños de polígonos de comparación (10 y 100 km<sup>2</sup>). Para tratar de ubicar un punto de quiebre, se modelaron las tendencias del PC (P-10:  $y = -9E-08x^2 + 0.0025x + 83.87$ , [R<sup>2</sup> = 0.5891]; P-100:  $y = -1E-07x^2 + 0.0035x + 79.358$ , [R<sup>2</sup> = 0.5891]), cuyo cruce definió un punto de coincidencia alrededor de 5 600 m ( $h$ ), lo cual se ubica dentro del rango de quiebre definido (Figura 6b). Finalmente, el valor de  $h$  estadísticamente más adecuado será aquel que más se aproxime a 5 600 m y que se ubique dentro del rango de 5 000 y 6 000 m. De acuerdo con esto, el proceso de definición de  $h$  que más se aproxima a este valor es el de distancia aleatoria media ( $h = 5\ 395$  m).

## Conclusiones

Es factible establecer un proceso estadístico de selección del ancho de banda ( $h$ ), para las estimaciones de densidad kernel de incendios forestales. De esta forma, se tiende a definir procesos y resultados replicables, comparables y compatibles, que no están condicionados a la participación de expertos, lo cual evita el uso de criterios subjetivos como la apreciación visual de las superficies continuas, que definen la variación espacial de la densidad de incendios forestales. Se encontró gran variación entre los valores del ancho de banda, lo cual implica que no existe un proceso único y universal para todos los casos. Esto debido a que las estimaciones de densidad kernel están condicionadas a varios factores, como la resolución espacial de la información (tamaño de celda), el ancho de banda, el número de puntos a considerar y las características intrínsecas del fenómeno en estudio.

### Fin de la versión en español

- Krisp, J. M., Peters, S., Murphy, C. E., & Fan, H. (2009). Visual bandwidth selection for kernel density maps. *ISPRS—Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 5, 445–454. doi: 10.1127/1432-8364/2009/0032
- Kulldorff, M., & Nagarwalla, N. (1995). Spatial disease clusters: Detection and inference. *Statistics in Medicine*, 14(8), 799–810. doi: 10.1002/sim.4780140809
- Kuter, S., Usul, N., & Kuter, N. (2011). Bandwidth determination for kernel density analysis of wildfire events at forest sub-district scale. *Ecological Modelling*, 222(17), 3033–3040. doi: 10.1016/j.ecolmodel.2011.06.006
- Mohammadi, F., Bavaghar, M. R., & Shabanian, N. (2014). Forest fire risk zone modeling using logistic regression and GIS: an Iranian case study. *Small-scale Forestry*, 13(1), 117–125. doi: 10.1007/s11842-013-9244-4
- Salvati, L., & Ferrara, A. (2015). Validation of MEDALUS fire risk index using forest fire statistics through a multivariate approach. *Ecological Indicators*, 48, 365–369. doi: 10.1016/j.ecolind.2014.08.027
- Seaman, D. E., & Powell, R. A. (1996). An evaluation of the accuracy of kernel density estimators for home range analysis. *Ecology*, 77(7), 2075–2085. Retrieved from <http://links.jstor.org/sici?sici=0012-9658%28199610%2977%3A7%3C2075%3AAEOTAO%3E2.0.CO%3B2-%23>
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. London, UK: Chapman & Hall/CRC. Retrieved from <https://ned.ipac.caltech.edu/level5/March02/Silverman/paper.pdf>
- Thompson, M. P., Calkin, D. E., Finney, M. A., Gebert, K. M., & Hand, M. S. (2013). A risk-based approach to wildland fire budgetary planning. *Forest Science*, 59(1), 63–77. doi: 10.5849/forsci.09-124.
- Turlach, B. A. (1999). *Bandwidth selection in kernel density estimation: A review*. Belgium: Institut de Statistique Université Catholique de Louvain. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.44.6770&rep=rep1&type=pdf>
- Vilar, H. L., Martín, I. M. P., & Martínez, V. F. J. (2011). Logistic regression models for human-caused wildfire risk estimation: analysing the effect of the spatial accuracy in fire occurrence data. *European Journal of Forest Research*, 130(6), 983–996. doi: 10.1007/s10342-011-0488-2
- Vrahimis, A. (2010). Smoothing methodology with applications to nonparametric statistics. Tesis Doctoral, University of Manchester. Retrieved from [https://www.research.manchester.ac.uk/portal/files/54503714/FULL\\_TEXT](https://www.research.manchester.ac.uk/portal/files/54503714/FULL_TEXT)