

quantil  
matemáticas aplicadas

# Validación de Modelos para la Predicción de Homicidios en Bogotá

## Seminario de Matemáticas Aplicadas Quantil



SECRETARÍA DE  
SEGURIDAD, CONVIVENCIA  
Y JUSTICIA



UNIVERSIDAD  
NACIONAL  
DE COLOMBIA

quantil

Febrero 24 de 2021

# Equipo Quantil

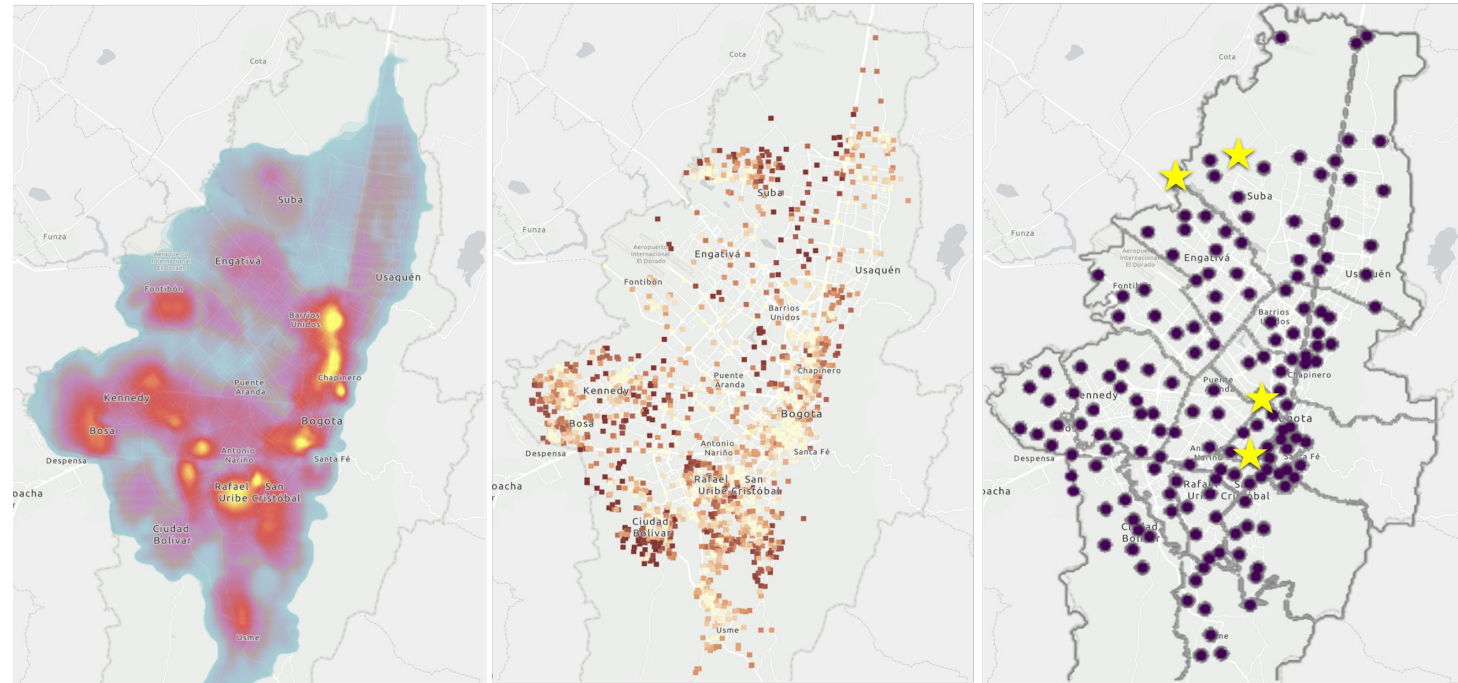


## Artículo de investigación

[Á. J. R. Villegas, J. S. M. Pabón, M. Dulce Rubio, S. Quintero, J. G. Vargas and H. García, "Spatio Temporal Sparsity in Homicide Prediction Models," in IEEE Access, vol. 10, pp. 14359-14367, 2022, doi: 10.1109/ACCESS.2022.3143858.](#)

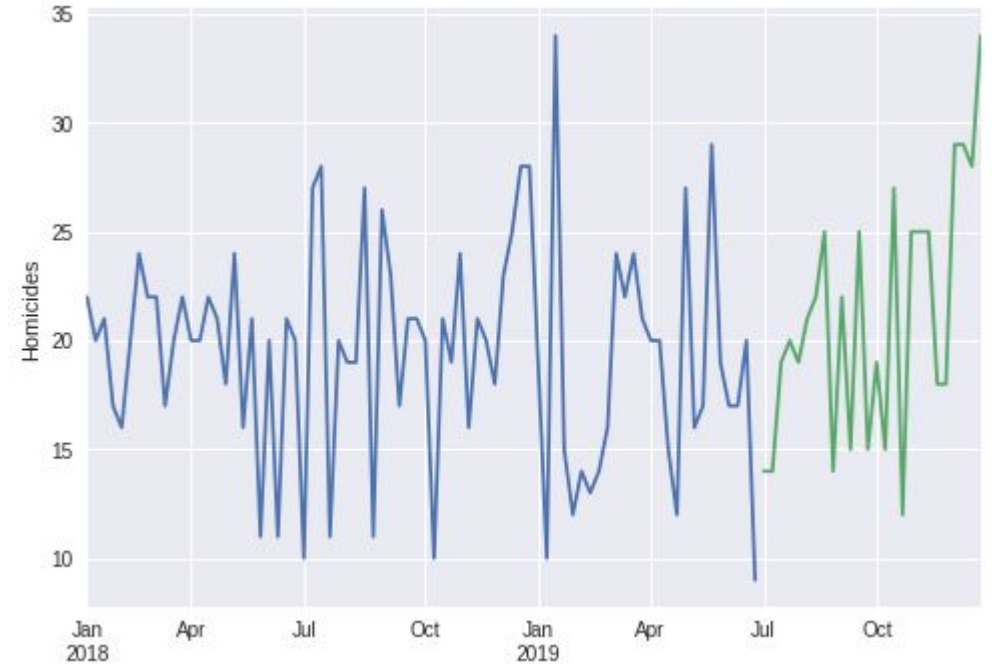
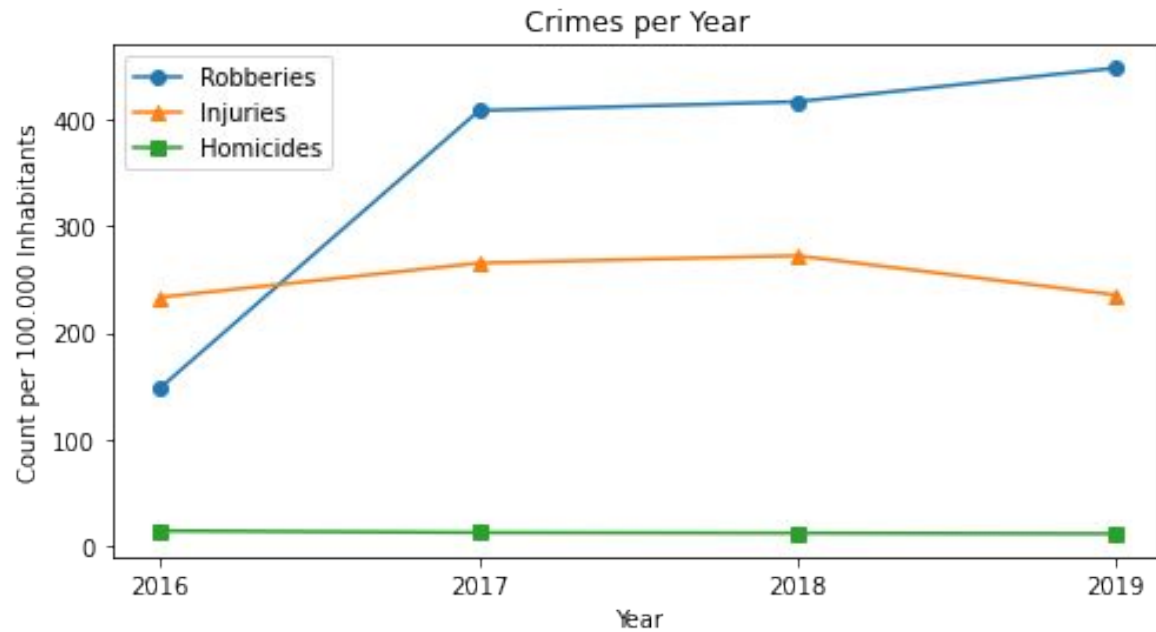
# Aplicaciones de los modelos

- » Uso eficiente de los recursos policiales
- » Monitoreo eficiente de cámaras
- » Ubicación eficiente de CAI



**Los homicidios  
son dispersos**

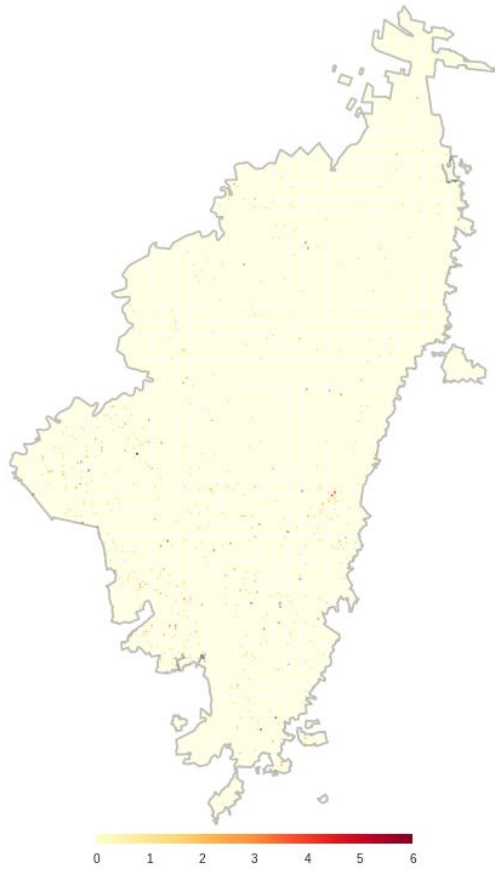
# Distribución temporal



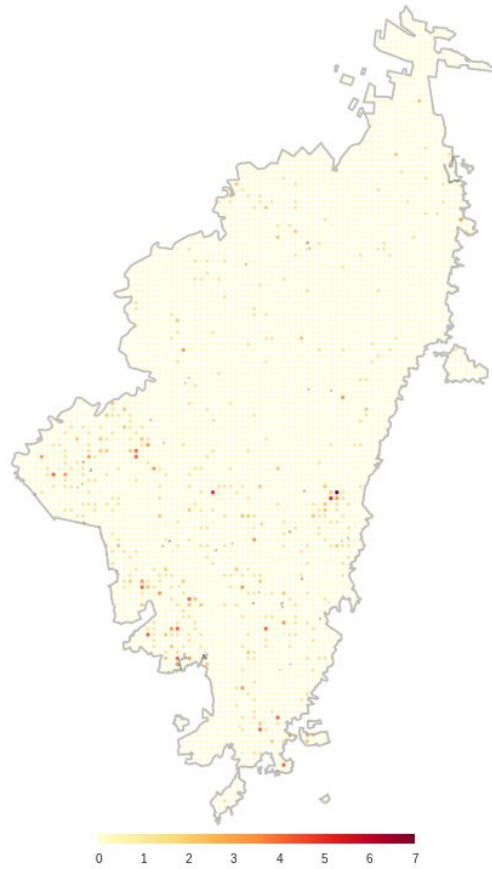
# Distribución espacial

Julio a diciembre de 2019

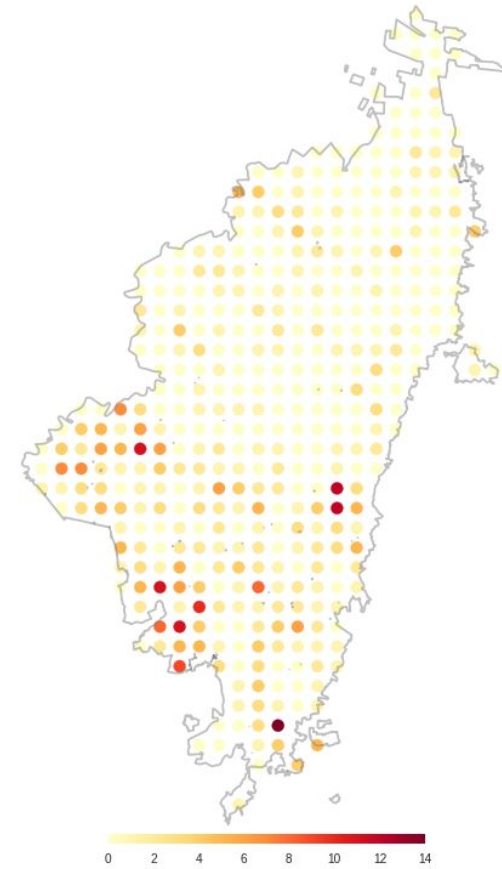
Resolution 150mts



Resolution 300mts



Resolution 1000mts



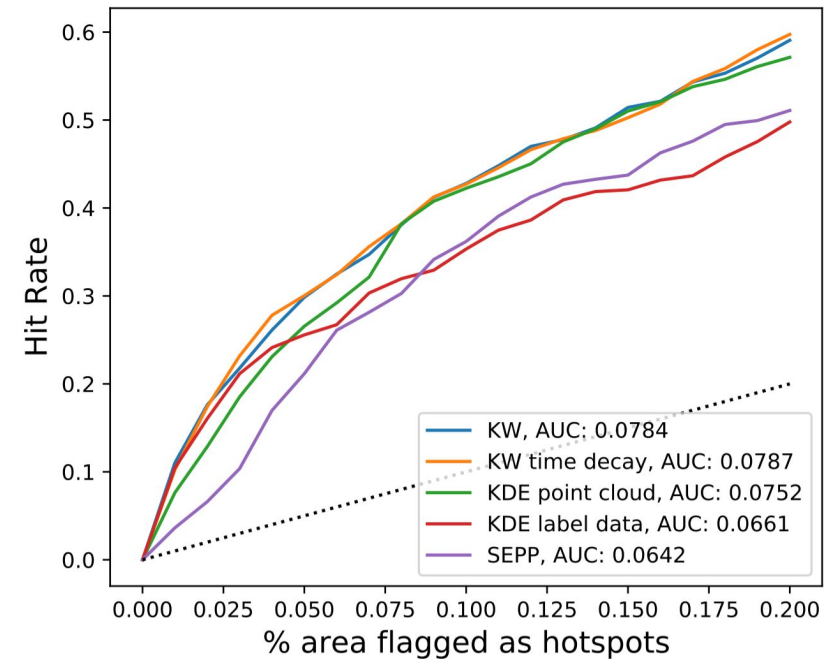
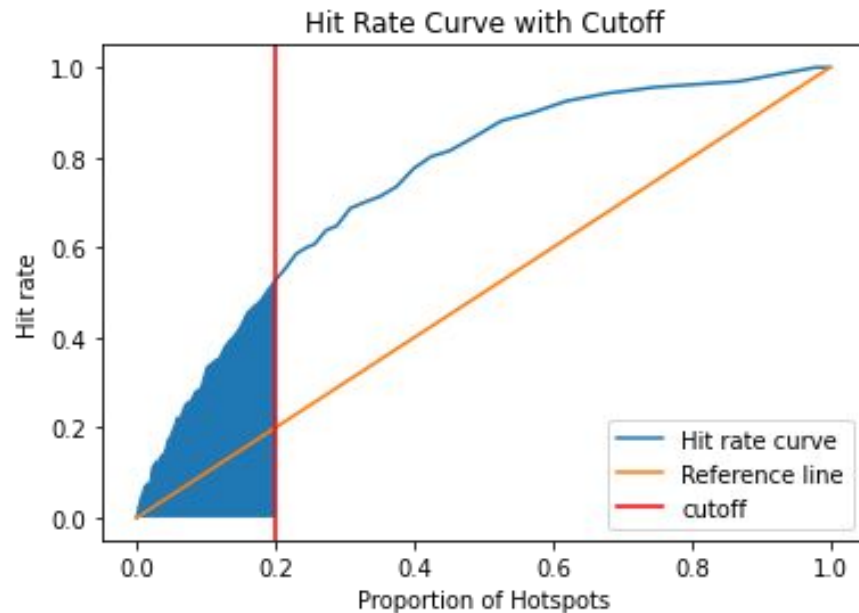
# **Métodos de Validación**



# Métricas - Hit Rate

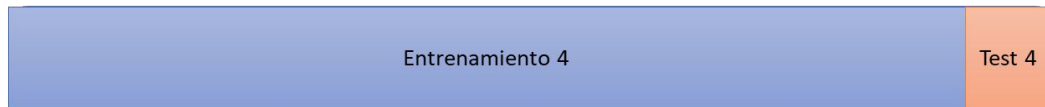
$$\text{Hit Rate} = \frac{\# \text{Eventos del conjunto de prueba en puntos calientes predichos}}{\# \text{Eventos en conjunto de prueba}}$$

$$\text{Porcentaje de área cubierta} = \frac{\text{Área total de los puntos calientes predichos}}{\text{Área total de estudio}}$$



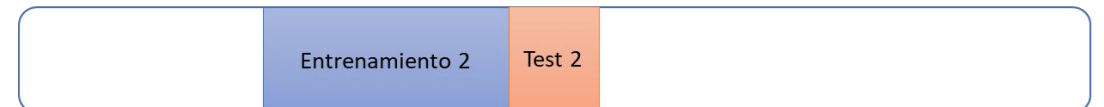
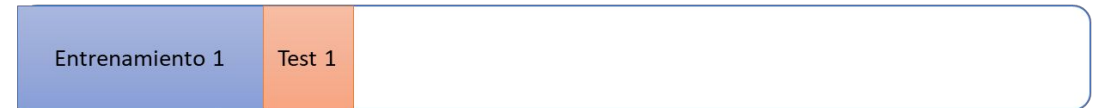
# Unidad temporal

Datos históricos de eventos criminales



tiempo

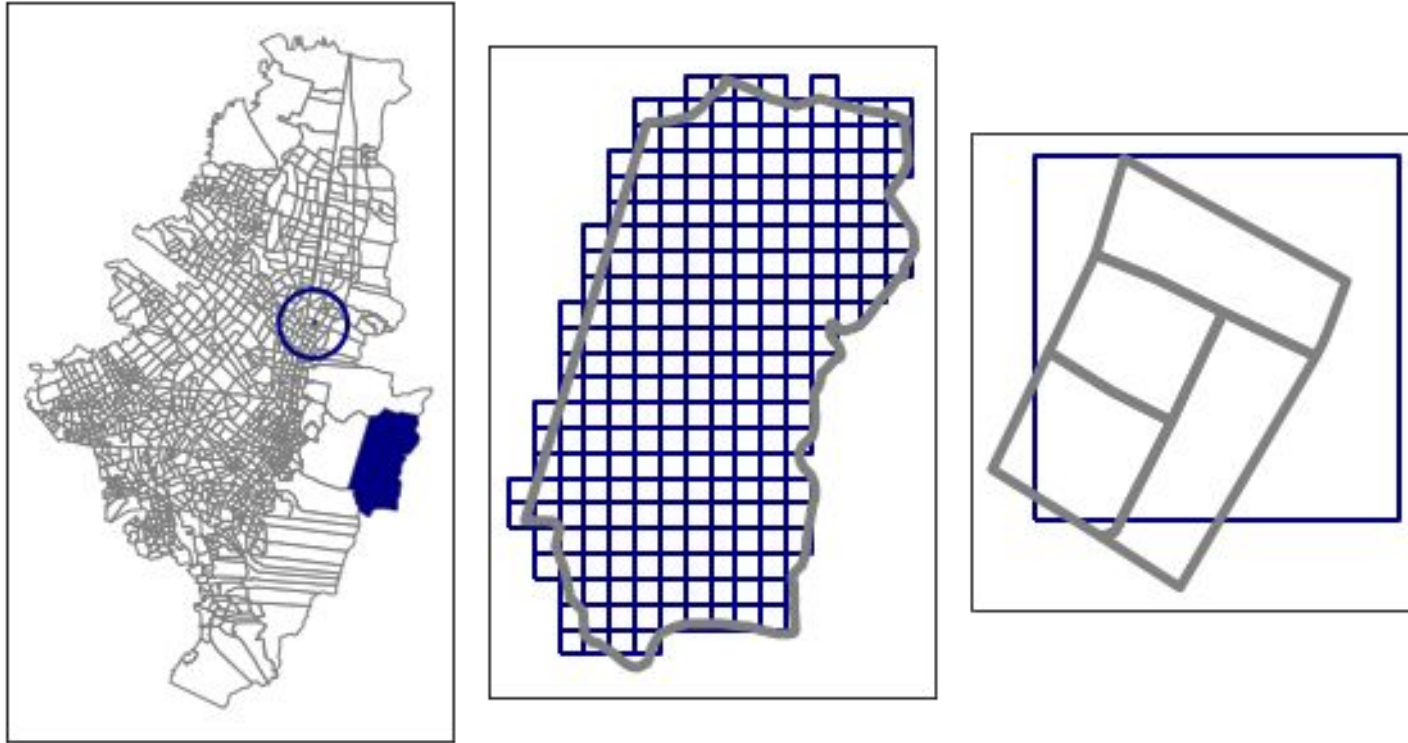
Datos históricos de eventos criminales



tiempo

Ventanas crecientes vs corredizas

# Unidad Espacial

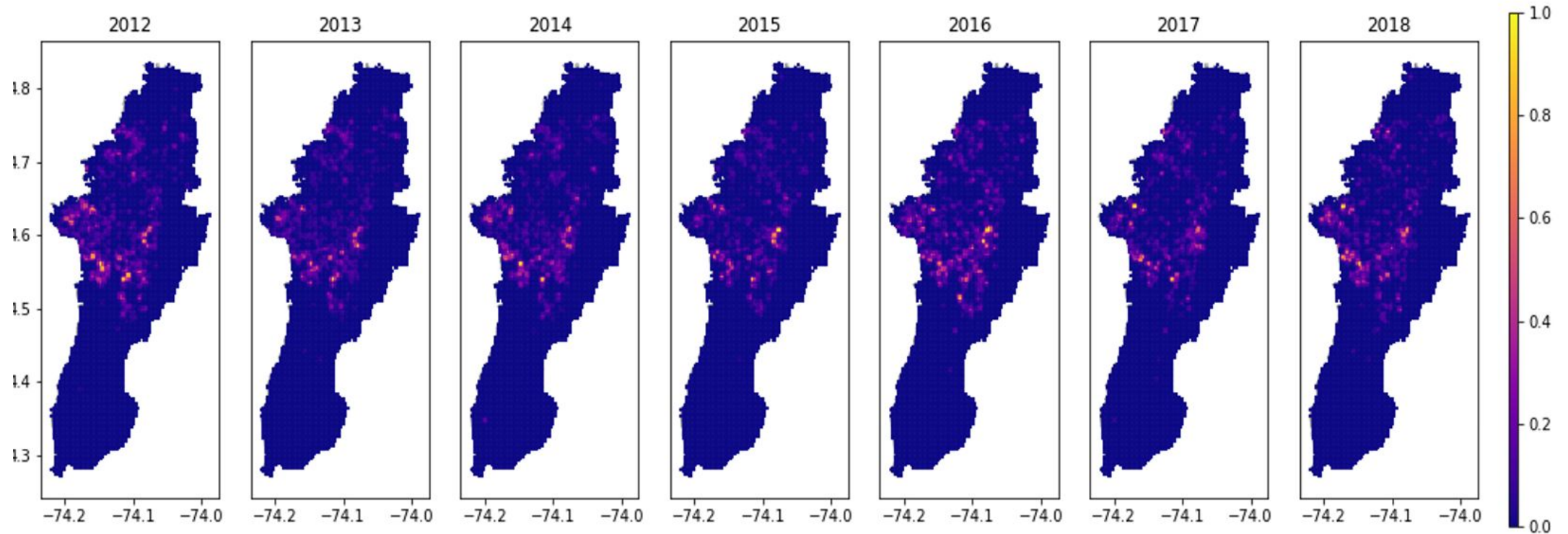


Cuadrantes vs Grilla regular

# **Modelos predictivos**

# Kernel de densidad

$$\lambda(x, y) = \sum_{t_i \in [T_1, T_2]} g(x - x_i, y - y_i).$$



# Procesos auto-excitantes

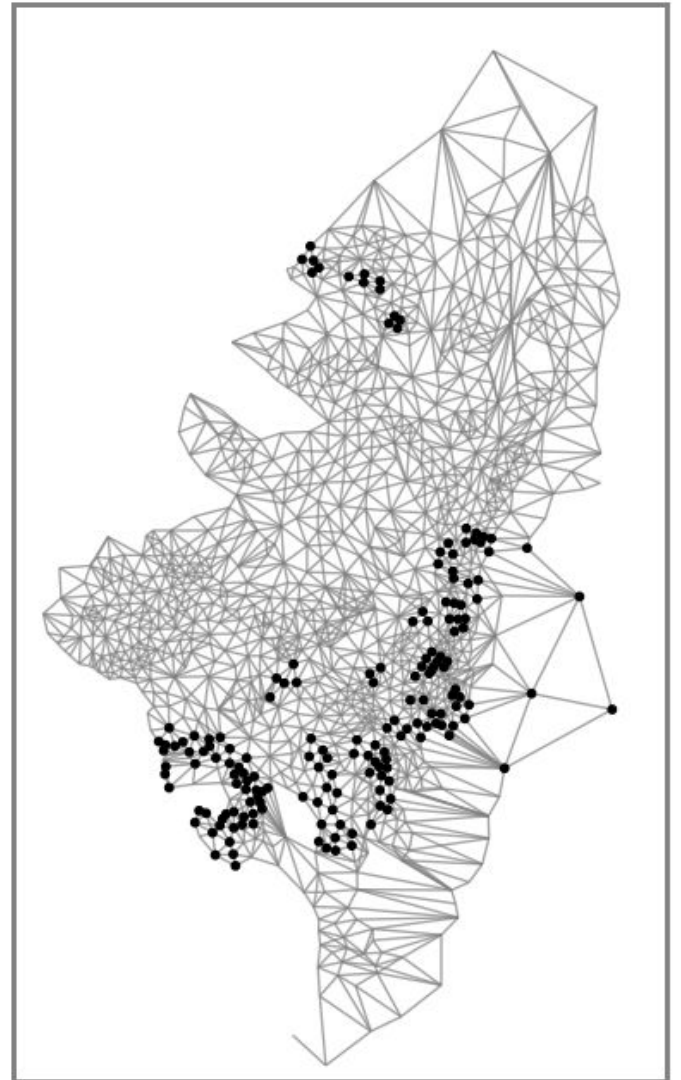
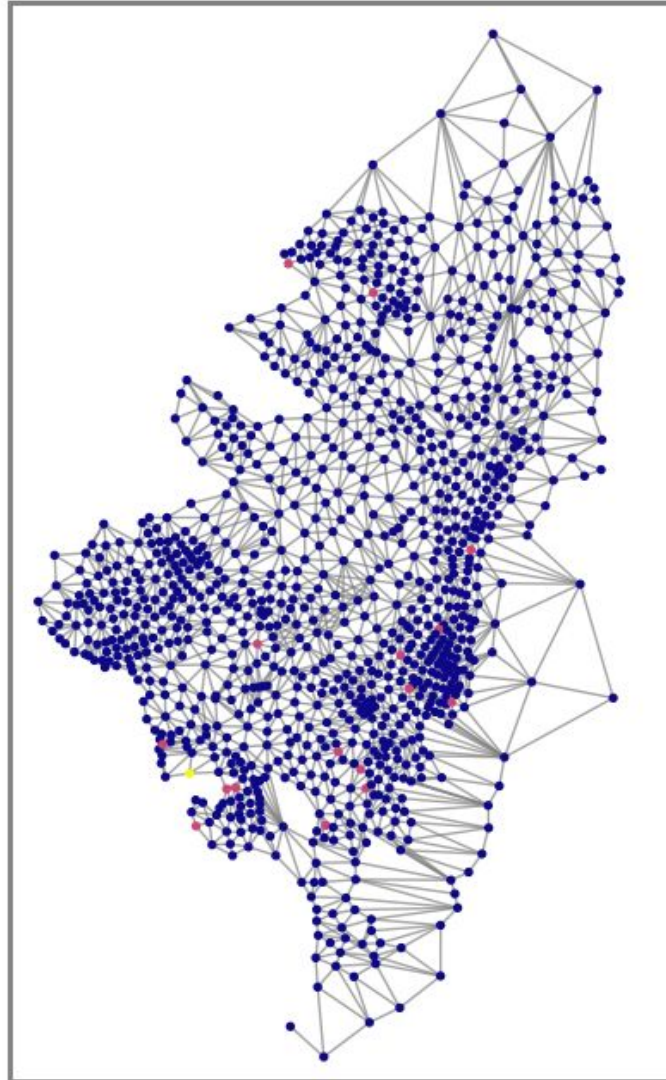
$$\lambda(x, y, t) = \mu(x, y) + \sum_{t > t_i} g(x - x_i, y - y_i, t - t_i, M_i).$$

- » Modelo propuesto por Mohler (2014).
- » Combina efectos persistentes con naturaleza auto-excitante.
- » Se enriquece con eventos correlacionados, en nuestro caso riñas.



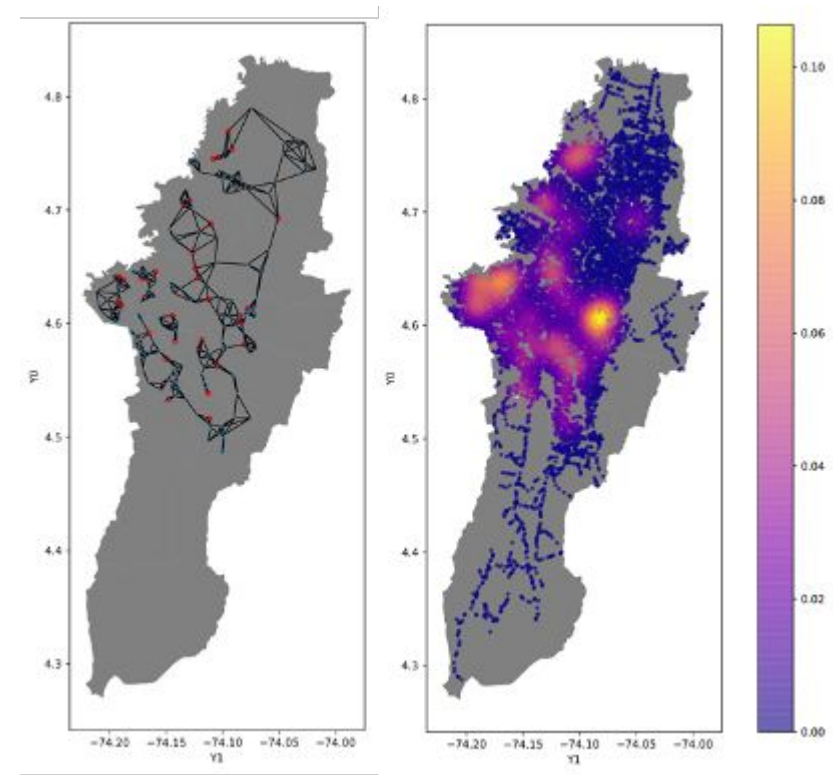
# GLOG: Graph Laplace of Gaussian

- » Detección de fronteras en la señal de homicidios sobre el grafo de (riñas en) la ciudad.
- » Regresión logística basada en las características de cada nodo.



# Kernel Warping

- » Marcamos las ubicaciones de homicidios y riñas.
- » Construimos grafo en base a distancias.
- » Consideramos ese grafo como variedad subyacente.
- » Calculamos matriz laplaciana del grafo.





# Deformación del kernel

Kernel usual

$$k(x, x_i) = \frac{1}{2\sigma^2} \exp\left(-\frac{\|x-x_i\|^2}{2\sigma^2}\right),$$

$$f(x) = \frac{1}{N} \sum_{i=1}^N k(x, x_i) = \frac{1}{N} \sum_{i=1}^N \frac{1}{2\sigma^2} \exp\left(-\frac{\|x-x_i\|^2}{2\sigma^2}\right).$$

Kernel deformado

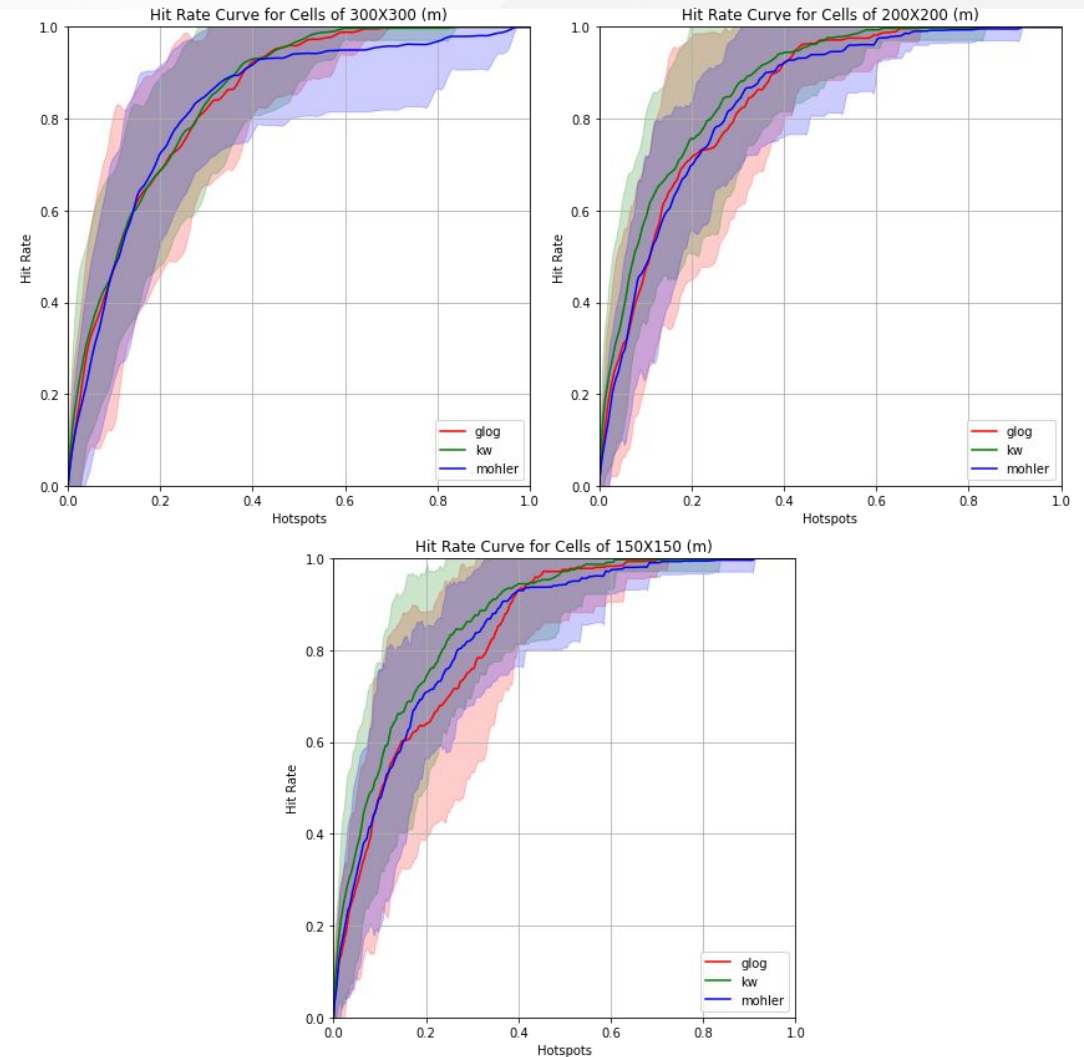
$$\tilde{k}_\sigma(x, s) = k_\sigma(x, s) - k_{xz}^T (I + \lambda K_{zz})^{-1} \lambda L k_{sz}, \quad (1)$$

$$\tilde{f}(x) = \frac{1}{|S|} \sum_i \tilde{k}_\sigma(x, s_i). \quad (2)$$

$$\hat{f}(x) = \frac{1}{|S|} \sum_i \exp(-\omega(t_x - t_{s_i})) * \tilde{k}_\sigma(x, s_i). \quad (3)$$

# Comparación de resultados

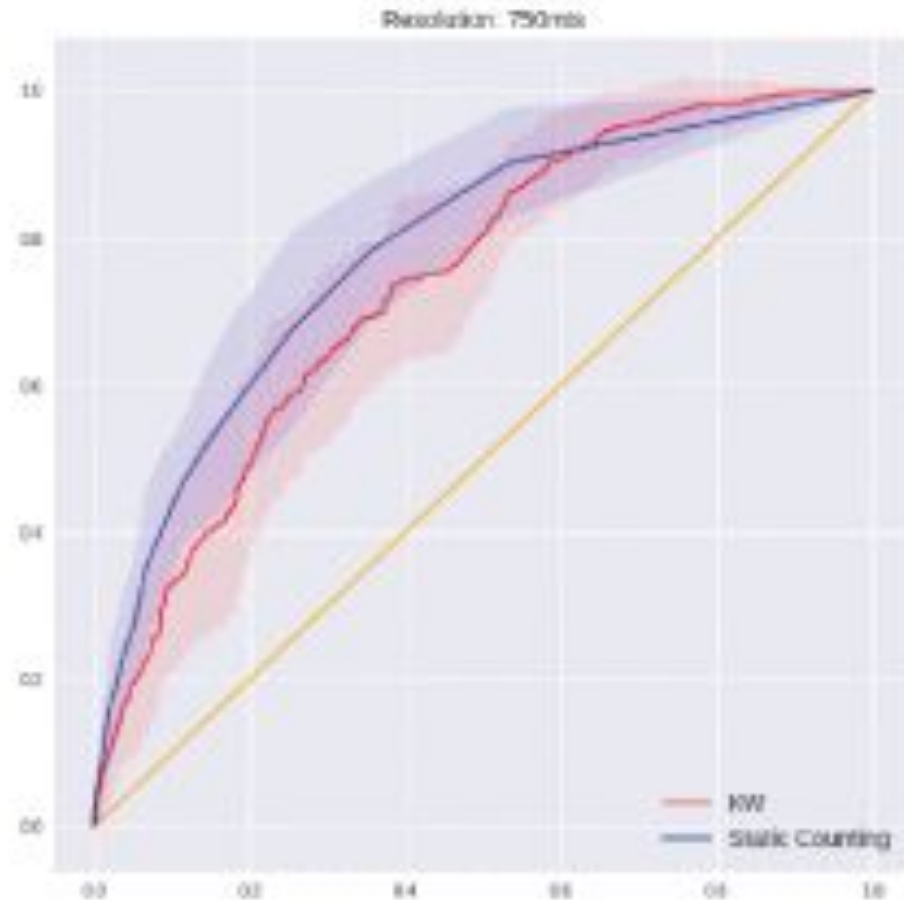
| Tamaño Celda (m) | Modelo | Hit Rate      | Hit Rate 5% ( $10^{-2}$ ) |
|------------------|--------|---------------|---------------------------|
| 300 × 300        | KW     | 0.773 (0.043) | 1.07 (0.48)               |
|                  | GR     | 0.843 (0.044) | 0.88 (0.37)               |
|                  | M      | 0.832 (0.056) | 0.76 (0.37)               |
| 200 × 200        | KW     | 0.795 (0.046) | 1.20 (0.46)               |
|                  | GR     | 0.844 (0.040) | 0.89 (0.39)               |
|                  | M      | 0.834 (0.057) | 0.83 (0.30)               |
| 150 × 150        | KW     | 0.791 (0.048) | 1.21 (0.45)               |
|                  | M      | 0.840 (0.040) | 0.93 (0.34)               |
|                  | GR     | 0.832 (0.043) | 0.77 (0.32)               |



# Modelo de Conteo

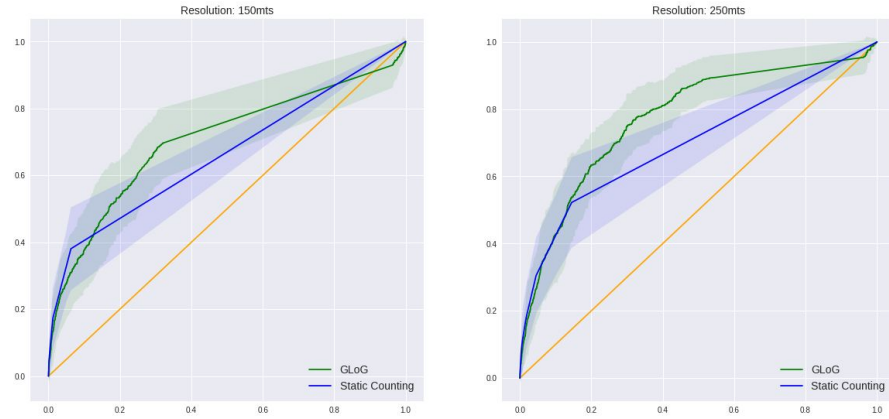
# Modelo de conteo

- » Para cada unidad espacial cuente los eventos durante el periodo de entrenamiento.
- » Variantes:
  - » Ventanas crecientes y corredizas.
  - » Factor de decaimiento temporal.
  - » Adicionar eventos relacionados

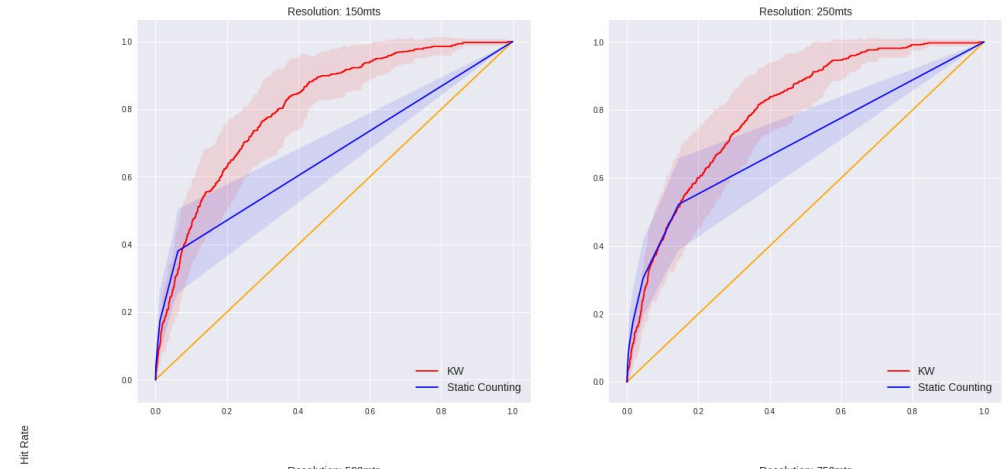


# Sensibilidad a la resolución

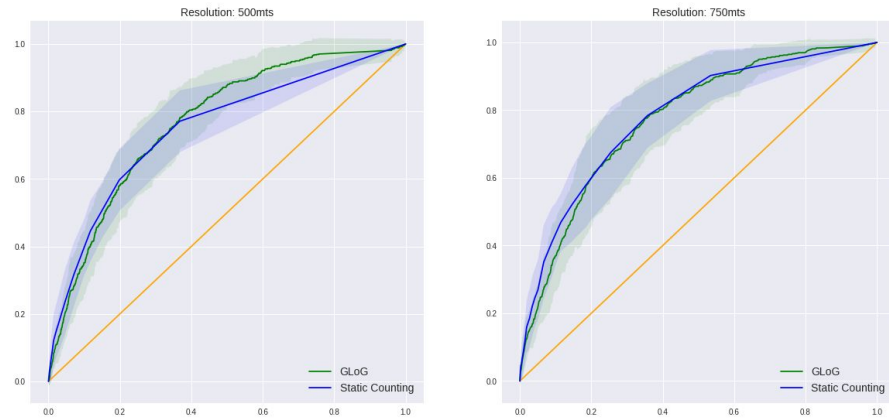
Hit Rate Curve: GLoG vs Static Counting



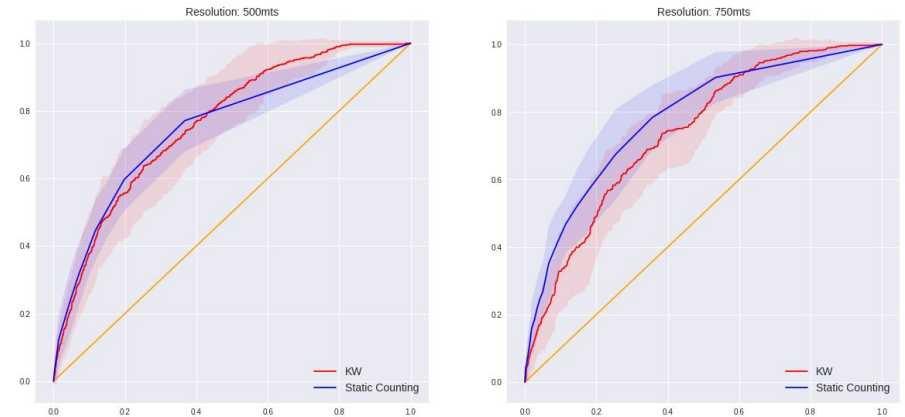
Hit Rate Curve: KW vs Static Counting



Hit Rate

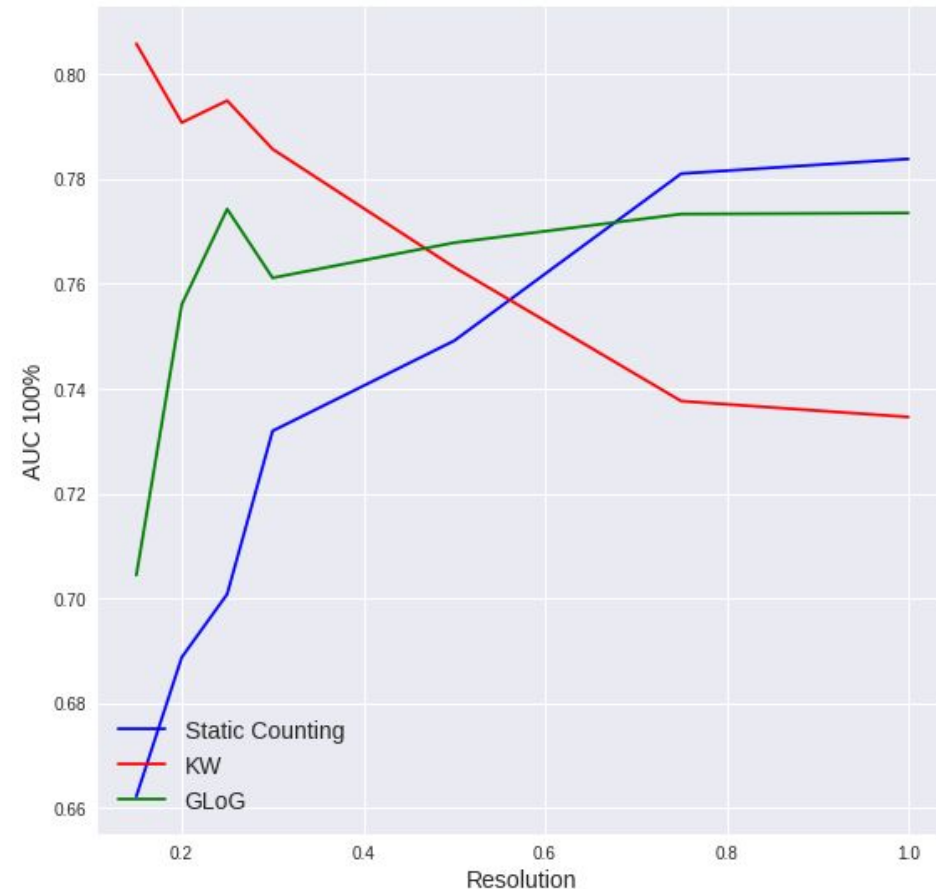


Hit Rate



# Sensibilidad a la resolución

- » Para celdas grandes el modelo de conteo tiene un desempeño superior.
- » Al aumentar la resolución los modelos sofisticados alcanzan y superan ese desempeño.

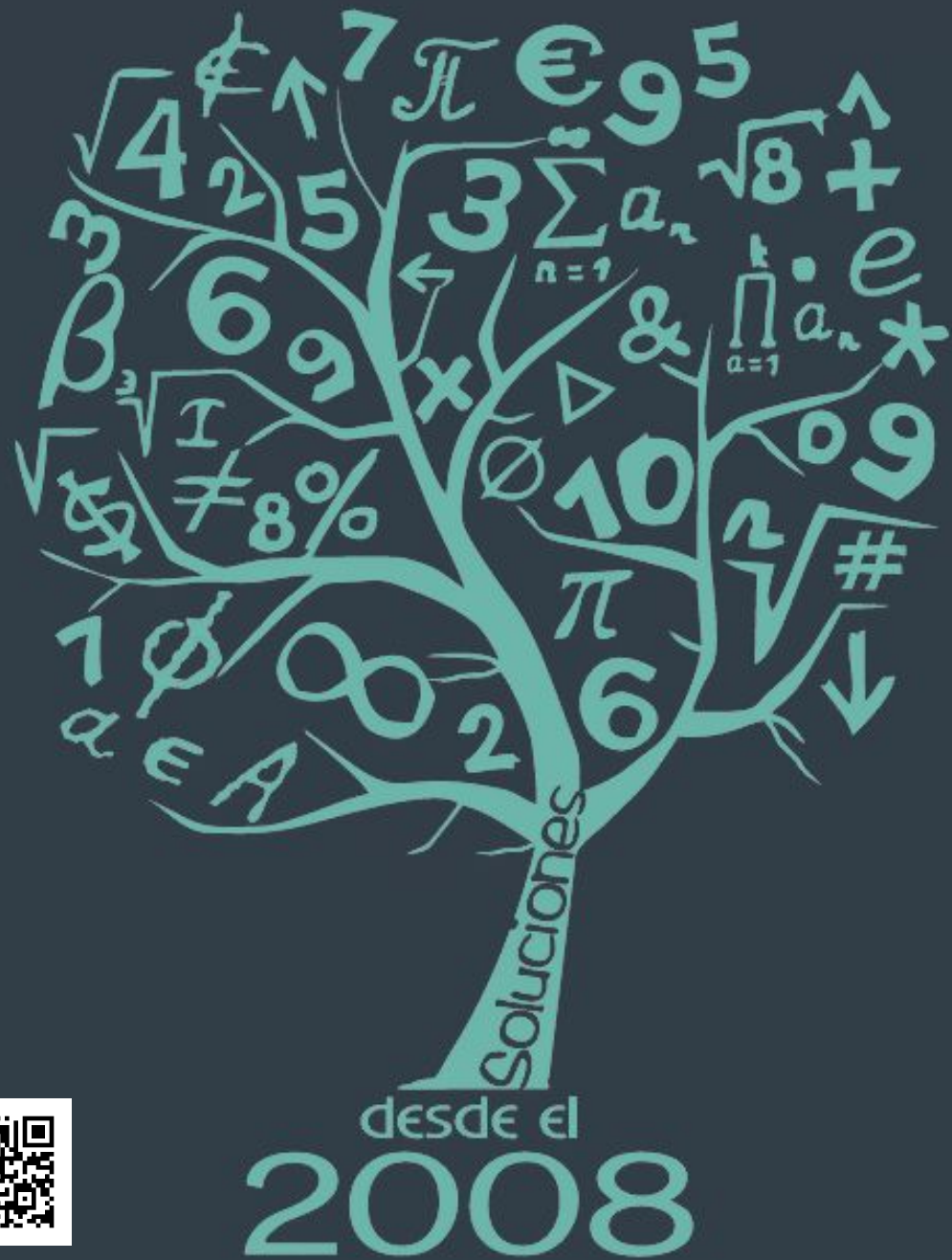




**GRACIAS**







quantil  
matemáticas aplicadas

## Conoce más de nosotros



[www.quantil.co](http://www.quantil.co)



[info@quantil.com.co](mailto:info@quantil.com.co)



+ 57 (1) 8051814



@QuantilMA