



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

CARACTERIZACIÓN Y PREDICCIÓN DE RIESGO CRIMINAL BASADO EN
KERNELS ESPACIO-TEMPORALES INCLUYENDO FENOMENOLOGÍA ESPACIAL

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIAS DE LA
INGENIERÍA, MENCIÓN ELÉCTRICA
MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL ELÉCTRICO

MARIO ALBERTO VERGARA LÓPEZ

PROFESOR GUÍA:
MARCOS ORCHARD CONCHA

MIEMBROS DE LA COMISIÓN:
JORGE SILVA SÁNCHEZ
MATÍAS ZAÑARTU SALAS

Este trabajo ha sido parcialmente financiado por CONICYT 3841/2016 y Carabineros de Chile

SANTIAGO DE CHILE
2017

RESUMEN DE LA MEMORIA PARA OPTAR AL
GRADO DE MAGÍSTER EN CIENCIAS DE LA INGENIERÍA, MENCIÓN ELÉCTRICA Y
TÍTULO DE INGENIERO CIVIL ELÉCTRICO
POR: MARIO ALBERTO VERGARA LÓPEZ
FECHA: 2017
PROF. GUÍA: SR. MARCOS ORCHARD CONCHA

CARACTERIZACIÓN Y PREDICCIÓN DE RIESGO CRIMINAL BASADO EN KERNELS ESPACIO-TEMPORALES INCLUYENDO FENOMENOLOGÍA ESPACIAL

La caracterización de riesgo criminal constituye un gran desafío, principalmente porque la ocurrencia de eventos criminales constituye un proceso puntual [10] de difícil predicción con lógicas similares a la de terremotos [1] y otros eventos geográficos [2]. En esta tesis, el estudio se acotó a la caracterización y predicción de riesgo criminal generado por Delitos de Mayor Connotación Social (DMCS) para los agrupados de Robo con Fuerza y Violencia en determinadas comunas de la Región Metropolitana. La elección de este tipo de delitos radicó en su naturaleza cuasi-periódica y con tendencia a la formación de *Hot-Spots*, lo que los vuelve idóneos para ser caracterizados mediante la teoría de kernels espacio-temporales [3]. El desafío principal fue generar y validar modelos descriptivos y predictivos del riesgo criminal a través de métricas *ad hoc* como AUC (*Area Under Curve*), curvas TIP (*True Incident Percentage*) y HRP (*High Risk Percentage*) [4] y la realización de estudios para determinar los factores más relevantes de la caracterización espacial considerando una caracterización temporal fija. La metodología implementada consta de cuatro pasos fundamentales: procesamiento de eventos criminales e información georeferenciada, generación de modelos espaciales que capten las tendencias históricas y recientes mediante modelos por mezcla de Gaussianas [5], adición de factores dinámicos mediante el entrenamiento de modelos temporales basándose en estudio de correlaciones y Wavelet [6] y finalmente la implementación de un modelo predictivo que se traduce en mapas de calor (*Hot-Spots*) para diferentes ventanas temporales. La inclusión de elementos espaciales como calles y el establecimiento de restricciones radicaron en mejorar el conocimiento del riesgo criminal con el objetivo de dar coherencia y sentido físico a los modelos espaciales en pos de mejorar el desempeño predictivo. Cabe mencionar que este trabajo de tesis se desprendió de un proyecto realizado en conjunto con CEAMOS (Centro de Modelamiento y Análisis en Seguridad) y Carabineros de Chile, donde los esfuerzos apuntaron a desarrollar un *software* funcional que computara predicciones en demanda para ubicar con anticipación recursos policiales. En ese sentido, el trabajo de tesis no solo apuntó a generar una metodología factible teóricamente sino que aplicable proveyendo soluciones eficientes y robustas para un amplio sector geográfico y diferentes ventanas de predicción. Los resultados demostraron que la metodología es capaz de caracterizar y predecir riesgo criminal para las comunas estudiadas, recalándose que en ciertas comunas la versión estática del algoritmo fue mejor que la dinámica. Por otro lado, mediante los estudios espaciales realizados, se establecen criterios para decidir qué parámetros de la caracterización espacial resultan ser óptimos en pos de un desempeño predictivo adecuado (cantidad de iteraciones del algoritmo de *Expectation Maximization*, cantidad de componentes óptimas, restricciones sobre valores propios de matrices de covarianza, entre otros parámetros). Finalmente, gracias a la estructura algorítmica desarrollada, la metodología en base a kernels espacio-temporales ofrece flexibilidad para agregar nuevos procedimientos espaciales como procesos Gaussianos [7], y procedimientos temporales que incluyan análisis espectrales más sofisticados.

Para todos aquellos que han puesto su granito de arena.

“So why should I want anything from you? The Court doesn't want anything from you. It receives you when you come, and it dismisses you when you go.”
The Trial - Franz Kafka

Agradecimientos

Agradezco a todas las personas que han formado parte de mi vida, quienes me han brindado herramientas para conocer el mundo y gracias a ello poder tomar las mejores decisiones que he podido. Con esto en mente, agradezco en primer lugar a mi familia: Haydée, Mario y Valeria, porque gracias a su cariño, enseñanza y apoyo incondicional, han forjado en mí la perseverancia y disciplina de poder alcanzar mis metas. Sin ellos todo hubiera sido más difícil, por lo que estoy agradecido de corazón.

A mi compañía tan *ad hoc*, Belén, gracias por todo.

Agradezco a todos mis amigos, a quienes he visto evolucionar y han visto en mí alguien en quien confiar: todas las experiencias que he vivido especialmente en el colegio, y la universidad las atesoro y son parte de mi constante aprendizaje. En particular quiero agradecer a Justin y Thamara, Franco y Fernando, Thiare.

A quienes hicieron posible este trabajo: gracias a todos y cada uno de los que conforman y conformaron CEAMOS, en especial a Mario Fernández, por sus valiosos consejos, no solo en lo profesional, sino también en la vida, a Leonardo por las excelentes discusiones teóricas y por supuesto a todo el equipo que trabajó duramente. A mi gran amigo y compañero: Pablo, porque nos las jugamos de inicio a fin concluyendo con éxito este proyecto, superando obstáculos y aprendiendo una enormidad sobre la marcha. A mi profesor guía Marcos Orchard: agradezco su confianza, sus brillantes ideas y entusiasmo en guiar este proceso. A Milena por su gran disposición, gestión y labor en todo momento.

Agradecer también a la comisión: Jorge Silva y Matías Zañartu por tener la disposición de leer este trabajo, corregirlo y de aportar con críticas constructivas. También a los que aportaron desde un inicio incluso antes de pensar hacer esta tesis en riesgo criminal: David, Francisco y Felipe T.

Agradezco a Carabineros de Chile, por su labor y gestión en la correcta realización de este proyecto. También a CONICYT 3841/2016 y FONDECYT 1140774 por financiar este trabajo y magister parcialmente.

Tabla de Contenido

1. Introducción	1
1.1. Motivación	1
1.2. Alcances	2
1.3. Objetivos	3
1.4. Hipótesis de Trabajo	4
1.5. Acuerdos de Confidencialidad	4
1.6. Estructura de la Tesis	5
2. Estado del Arte y Marco Teórico	6
2.1. Metodologías para Estimación de Riesgo Criminal	6
2.2. Introducción a Procesos Puntuales	11
2.3. Definición de Procesos Puntuales	12
2.4. Función de Intensidad	13
2.5. Estimación de Función de Intensidad	13
2.5.1. Definición de Kernels y Tipos	13
2.5.2. <i>Kernel Density Estimation</i>	14
2.5.3. Modelo por Mezcla de Gaussianas	16
2.5.4. Métricas para Análisis de Modelos Espaciales	16
2.5.5. Métricas de Validación Predictiva	21
3. Metodología de Caracterización de Riesgo Criminal Incluyendo Elementos Fenomenológicos	23
3.1. Procesamiento de Datos	24
3.1.1. Separación de Datos	24
3.2. Procesamiento Espacial	25
3.2.1. Elección Previa de Parámetros	25
3.2.2. Restricciones Paramétricas	28
3.3. Procesamiento Temporal	30
3.3.1. Preprocesamiento de Series Temporales	31
3.3.2. Modelo por Mezcla de Kernels Periódicos	33
3.3.3. Reconstrucción de Señales Mediante Wavelet	34
3.4. Procedimiento de predicción	37
3.4.1. Proyección de Señales Pretéritas	38
3.5. Resumen del Algoritmo a Utilizar	38
4. Caso de Estudio: Modelación de Riesgo Criminal Espacio-Temporal con	

Datos Reales de Comunas del Gran Santiago	40
4.1. Validación de Metodología Utilizando una Comuna	41
4.1.1. Descripción Zona de Estudio	41
4.1.2. Base de Datos	41
4.1.3. Resultados Preliminares	43
4.1.4. Estudio de Criterios de Término de Iteraciones en Modelos Espaciales	50
4.1.5. Estudio Cantidad de Componentes Iniciales en Modelos Espaciales . .	51
4.1.6. Estudio de Mecanismo sobre Valores Propios de Matrices de Covarianza	56
4.1.7. Estudio de Inclusión de Calles en la Caracterización de Riesgo	59
4.1.8. Resultados Predictivos Generales	62
4.2. Resultados de Metodología para Diferentes Comunas	69
4.2.1. Resultados AUC Promedio para Cada Comuna de Estudio	70
4.2.2. Discusión	71
4.3. Comparación de Resultados Predictivos con Otros Métodos	72
5. Conclusiones	76
5.1. Trabajo Futuro	78
5.1.1. Con Respecto a Métricas de Evaluación Espaciales	78
5.1.2. Propuesta de Método <i>Online</i> para Determinar Número de Componentes Óptimos para GMM	78
5.1.3. Con Respecto a Condiciones Iniciales para GMM	78
5.1.4. Con Respecto a Mecanismo sobre Valores Propios de Matrices de Covarianza	79
5.1.5. Con Respecto a Algoritmo EM	79
5.1.6. <i>Framework</i> de Entrenamiento de Parámetros	79
Bibliografía	81

Índice de Tablas

4.1. Muestra de nombre de delitos y código de delitos para agrupado alcoholes.	42
4.2. Cantidad de delitos según agrupados anual	42
4.3. Parámetros para caracterización <i>a priori</i> y <i>a posteriori</i> típicas, donde <i>n_components</i> es la cantidad de componentes de la GMM, <i>covariance_type</i> indica el tipo de covarianza siendo en este caso con vectores propios arbitrarios, <i>min_covar</i> es el mínimo valor de la covarianza diagonal, <i>params</i> corresponde a los parámetros a entrenar siendo <i>weights</i> , <i>means</i> y <i>covariances</i> y <i>niter</i> es la cantidad de iteraciones del algoritmo EM.	44
4.4. Resumen mejores λ análisis estático fechas 2015-01-01 a 2015-05-01.	64
4.5. Resumen mejores λ análisis estático fechas 2016-01-01 a 2016-05-01.	65
4.6. Rangos de AUC de predicción para años 2015 y 2016 del análisis estático	65
4.7. Resumen mejores λ análisis series de tiempo perfectas fechas 2015-01-01 a 2015-05-01.	66
4.8. Resumen mejores λ análisis series de tiempo perfectas fechas 2016-01-01 a 2016-05-01.	66
4.9. Rangos de AUC de predicción para años 2015 y 2016 del análisis con series de tiempo perfectas	66
4.10. Resumen mejores λ análisis predicciones estándar fechas 2015-01-01 a 2015-05-01.	67
4.11. Resumen mejores λ análisis series de tiempo estándar fechas 2016-01-01 a 2016-05-01.	68
4.12. Rangos de AUC de predicción para años 2015 y 2016 del análisis con series de tiempo estándar	68
4.13. Porcentaje de aciertos para un 10% de cobertura para las diferentes comunas trabajadas considerando el mejor resultado según el turno policial y versión del algoritmo (estático o dinámico). Basado en los resultados de Anexos: Resultados AUC promedio para cada comuna de estudio.	75
5.1. Nombre de delitos y sus agrupados según el tipo DMCS.	85

Índice de Ilustraciones

2.1.	Métodos matemáticos para predecir ocurrencia de crímenes de acuerdo al problema que se desea solucionar. Fuente: Libro de W. Perry <i>et al.</i> [8].	7
2.2.	Diferentes métodos para la construcción de <i>Hot-Spots</i> . a) ocurrencia de eventos criminales, b) estimación de elipses mediante la desviación estándar de los delitos a su centro, c) designación de riesgo según zonas administrativas policiales, d) mapeo del territorio a través de grillas y finalmente en e) la estimación del riesgo mediante KDE (<i>Kernel Density Estimation</i>). Fuente: S. Chainey <i>et al.</i> [9].	8
2.3.	Estimación de las intensidades de fondo para el caso temporal y espacial. Fuente: [10].	9
2.4.	Kernel periódico con $x \in [0, 2]$, $x' = 0.5$, $\sigma = 1.0$, $p = 0.7$ y $l = [0.25, 0.35, 0.5]$	15
2.5.	Coefficiente de Silhouette para cada dato de entrenamiento considerando datos artificiales generados por 4 clusters versus un clustering mediante 10 clusters.	18
2.6.	Promedio del coeficiente de Silhouette para diferente cantidad de clusters o componentes del algoritmo GMM.	18
2.7.	Métrica de desviaciones para diferente número de clusters o componentes del algoritmo GMM.	19
3.1.	Metodología general para caracterización y predicción de riesgo criminal. . .	23
3.2.	Esquema general para procesamiento espacial.	26
3.3.	Ejemplificación de mecanismo	30
3.4.	Esquema general para procesamiento temporal.	30
3.5.	Método de correlaciones: para un conjunto de centros \mathbf{c} y periodos \mathbf{p} finitos se calcula la correlación entre la serie temporal suavizada $\bar{R}(t)$ con el kernel periódico para un determinado centro y periodo (construcción del conjunto $\Omega_{\text{corrj}}^{\text{pr}}$). Finalmente se ordenan según su grado de correlación y se escogen los elementos más correlacionados (construcción del conjunto $\mathcal{PC}_{\text{corrj}}^{\text{pr}}$). Fuente: tesis del Sr. Pablo Fuentes [6].	35
3.6.	Ecuaciones (izquierda) que definen funciones bases típicas de Transformada Wavelet Continua con su respectiva forma (derecha). Fuente: Trabajo de C. Torrence <i>et al.</i> [11]	36
3.7.	Señal temporal típica suavizada asociada a un kernel espacial con su respectivo EEW para periodos de tiempo de 0 a 120 días, y su histograma de energía promedio para bandas de ancho de 3 días. Fuente: tesis del Sr. Pablo Fuentes [6].	36
3.8.	Resumen de metodología	39

4.1.	División política del Gran Santiago y comuna de Puente Alto en color celeste.	41
4.2.	Caracterización espacial de modelos.	45
4.3.	Pertenencia de eventos \mathcal{E}_{pr} a $\mathcal{K}_j^{pr} \forall j = 1, \dots, 179$ (considerando centroides en color rojo) y construcción de histograma de cantidad de eventos de cada \mathcal{K}_j^{pr} .	45
4.4.	Serie de tiempo de ejemplo suavizada con KDE con su respectiva matriz de correlaciones destacando los primeros 10 pares de centros y periodos (en unidades de 8 horas)	46
4.5.	Serie de tiempo suavizada con KDE de un kernel de \mathcal{K}_j^{pr} y su versión reconstruida mediante Transformada Wavelet (ver Sección 3.3.3).	47
4.6.	Predicciones de 1 a 4 pasos considerando un 30 % del área de cobertura total de la comuna.	48
4.7.	Grillas de estudio para comuna de Puente Alto.	48
4.8.	Curva HRP v/s TIP para 4 pasos de predicción con el respectivo valor AUC y un ejemplo para 1 paso de predicción mostrando 164 grillas de mayor riesgo (3.9 %) con acierto de 3 de 7 delitos (42.8 %).	49
4.9.	Iteraciones de modelos espaciales <i>a priori</i> y <i>a posteriori</i>	51
4.10.	Criterios multi-objetivo (<i>Silhouette</i> con Aciertos y <i>Silhouette</i> con Desviaciones)	52
4.11.	Relación entre criterio de Desviaciones (Cohesión) y Aciertos para modelo <i>a priori</i> y <i>a posteriori</i> considerando 10 a 300 componentes espaciales correspondiente a los puntos que parten desde el borde inferior izquierdo y terminan en el borde superior derecho.	54
4.12.	Valores de AUC a un paso y promedio de tres pasos para diferentes componentes espaciales de modelo <i>a priori</i> y <i>a posteriori</i>	55
4.13.	Elipse de cada componente de modelo <i>a posteriori</i> graficando cada Gaussiana con 0.9 de probabilidad. El mecanismo se ejecuta considerando $r^2 = 6$ y $\gamma = 0.5$.	56
4.14.	Métricas espaciales para modelo <i>a posteriori</i> de 40 componentes iniciales, utilizando el mecanismo para diferentes γ y diferentes r^2 tal que N_{rest} (cantidad de componentes afectadas por el mecanismo) cambie uno en uno. Obs: Los criterios están normalizados entre 0 y 1.	57
4.15.	Valores de AUC a un paso y promedio de tres pasos para diferentes componentes espaciales de modelo <i>a priori</i> y <i>a posteriori</i>	59
4.16.	Gráfico de calles de la comuna de estudio diferenciando el valor de riesgo de cada una tanto para modelo <i>a priori</i> como <i>a posteriori</i> . Para este ejemplo las calles de importancia se grafican en escala de colores a partir del percentil 80 de riesgo, y las que están bajo este umbral son pintadas azul.	60
4.17.	Curva HRP vs TIP para modelo con geometría de calles para la fecha 2016-05-03.	61
4.18.	Resultados de algoritmo con segmentos de calles para tres pasos de predicción para la fecha 2016-05-03 mostrando el 20 % de los segmentos de mayor riesgo y en puntos rojos los delitos de predicción de la fecha.	62
4.19.	Valor promedio de AUC de un 1 a 3 pasos y cantidad de veces que AUC fue máximo para diferentes valores de λ para las fechas 2015-01-01 a 2015-05-01.	64
4.20.	Valor promedio de AUC de un 1 a 3 pasos y cantidad de veces que AUC fue máximo para diferentes valores de λ para las fechas 2016-01-01 a 2016-05-01.	64
4.21.	Valor promedio de AUC de un 1 a 3 pasos y cantidad de veces que AUC fue máximo para diferentes valores de λ para las fechas 2015-01-01 a 2015-05-01.	65
4.22.	Valor promedio de AUC de un 1 a 3 pasos y cantidad de veces que AUC fue máximo para diferentes valores de λ para las fechas 2016-01-01 a 2016-05-01.	66

4.23. Valor promedio de AUC de un 1 a 3 pasos y cantidad de veces que AUC fue máximo para diferentes valores de λ para las fechas 2015-01-01 a 2015-05-01.	67
4.24. Valor promedio de AUC de un 1 a 3 pasos y cantidad de veces que AUC fue máximo para diferentes valores de λ para las fechas 2016-01-01 a 2016-05-01.	68
4.25. Comunas de estudio de la Región Metropolitana.	69
4.26. AUC's promedio de 1 a 3 pasos de predicción considerando versión estándar , perfecta y estática del algoritmo en el intervalo de fechas 2016-05-01 2016-08-01 para comunas de estudio.	71
4.27. Resultados predictivos diarios usando celdas de 200x200[m ²]. En cuanto al método prospectivo los parámetros de kernels son 400 [m] espaciales y 8 [semanas] (figura de la izquierda) y parámetros óptimos (figura de la derecha) donde se considera 200 [m] espaciales y 39 [semanas] temporales. La intensidad de fondo para el método de procesos puntuales corresponde a un ancho de banda de 300 [metros] (figura de la izquierda) seleccionado mediante validación cruzada y 130 [m] (figura de la derecha) seleccionado para optimizar el número de eventos criminales a predecir. Fuente: Trabajo de Mohler <i>et al.</i> [10]	73
4.28. AUC para eventos criminales de Charlottesville desde Marzo 2014 a Febrero 2005 (izquierda). Curva HRP/TIP para mes de Julio 2004 (derecha). En estos resultados se compara los algoritmos propuestos por Wang <i>et al.</i> consistentes en modelos aditivos GAM (línea verde y azul) con el típico modelo mediante Hot-spots (línea negra) y otro llamado <i>Generalized Linear Models</i> (GLM) (línea roja). Fuente: Trabajo de Wang <i>et al.</i> [4]	74
5.1. <i>Framework</i> de entrenamiento sistemático de parámetros.	80
5.2. AUC's promedio de 1 a 3 pasos de predicción para intervalo de fechas 2016-05-01 2016-08-01 para comunas de estudio.	88
5.3. AUC's promedio de 1 a 3 pasos de predicción para intervalo de fechas 2016-05-01 2016-08-01 para comunas de estudio.	89
5.4. AUC's promedio de 1 a 3 pasos de predicción para intervalo de fechas 2016-05-01 2016-08-01 para comunas de estudio considerando las tres versiones del algoritmo.	90
5.5. AUC's promedio de 1 a 3 pasos de predicción para intervalo de fechas 2016-05-01 2016-08-01 para comunas de estudio considerando las tres versiones del algoritmo.	91

Capítulo 1

Introducción

1.1. Motivación

La caracterización de riesgo criminal constituye un gran desafío tanto desde el punto de vista criminológico como matemático, principalmente por la dificultad que conlleva comprender la fenomenología del crimen, desentrañar los factores que lo originan y cómo es posible manejarlo. Es por eso que abarca diversas disciplinas tales como sociología, antropología, psicología entre otras las cuales se enfocan comprender la naturaleza del crimen, como también la matemática, estadística y geografía encargadas de generar modelos matemáticos para caracterizar los fenómenos que subyacen de ésta. El esfuerzo por coordinar cada una de las disciplinas, validar sus hipótesis y hacer de los modelos herramientas efectivas para el quehacer de las fuerzas policiales es complejo, abarcando largos plazos tanto en investigación como de experimentación.

Es por esto que las fuerzas policiales a nivel global trabajan en conjunto con universidades y centros de investigación para generar predicciones de eventos criminales con el objetivo de ubicar los escasos recursos policiales de manera eficiente, y así evitar la ocurrencia de eventos criminales. Actualmente existen variadas soluciones a nivel mundial correspondiente a *software* que procesan datos y generan predicciones para las fuerzas policiales. Predpol[12] es uno de estos *software* que creció como un proyecto de investigación entre el Departamento de Policía de Los Angeles y UCLA en California con el objetivo de anticipar posibles eventos criminales en ventanas espacio-temporales (cajas espaciales de 150x150[m] y para diferentes turnos policiales durante el día), que actualmente se posiciona como solución en varios países. Asimismo en Milán se tiene *Transcrime*, y en Memphis, mediante la colaboración de IBM[13] se tiene un *software* de predicción. Prestigiosas universidades como la UCL[14] en Londres, ICURS y PIMS en Canadá, tienen vasta experiencia en investigación e implementación de métodos matemáticos predictivos.

Chile no está exento de esta tendencia mundial, debido a que la delincuencia es una de las principales preocupaciones de las autoridades de cada país. Según el Balance de la Delincuencia en Chile del año 2015 publicado por la Fundación Paz Ciudadana [15], la delincuencia se mantiene dentro de las tres prioridades de la ciudadanía, más aún según el Centro de

Estudios Públicos para el 48 % de las personas sigue siendo uno de los problemas donde el gobierno debiese dedicar más esfuerzo en solucionar [16]. En el mismo balance se menciona además que la mayoría de delitos como hurtos, robos e intentos de robos y homicidios se concentran en espacio público y donde además los delitos con violencia han aumentado significativamente. Por su parte resultados de la XII Encuesta Nacional Urbana de Seguridad Ciudadana (ENUSC) [17], que mide la percepción de inseguridad de la población referente al año 2015, señala una tendencia al alza de la delincuencia en el país desde el año 2012.

Dado lo anterior, la delincuencia resulta ser un tema central, y en este sentido, los Delitos de Mayor Connotación Social (DMCS) específicamente los “Robo con Fuerza” y “Robo con Intimidación” son protagonistas. Según ENUSC, un 4,6 % de la muestra declara haber sido víctima de robo con violencia o intimidación en los últimos 12 meses, lo que suma más de 200 mil casos. En cuanto a Carabineros de Chile, en su Informe Anual 2015[18] cuantifica aproximadamente 300 mil denuncias de delitos tipificados como “Contra las Personas”, de las que 41.403 corresponden a robo con intimidación y 17.712 a robo con violencia. Totalizando se tiene que entre ambos tipos de delitos corresponden al 20,92 % de las denuncias totales del ítem mencionado, equiparándose a la cifra del año 2014 que fue de un 19,34 %.

Ante la creciente alza de los DMCS y delitos en vía pública, Carabineros de Chile implementó una plataforma llamada SAIT (Sistema de Análisis de Información Territorial)[19] cuyo objetivo en sus inicios fue meramente realizar análisis territoriales, hoy en día constituye una herramienta para planificar y mediante el conocimiento de expertos dar cuenta de las zonas de mayor riesgo criminal y así planificar en conjunto con las patrullas policiales rutas preventivas. Si bien la tarea de identificar zonas de riesgo con mapas de calor mediante ventanas móviles de datos y ubicar recursos policiales se realiza, actualmente no se posee un *software* que sea capaz de generar predicciones espacio-temporales y muestre el comportamiento dinámico de *Hot-Spots* criminales con una resolución temporal para cada turno policial. Tampoco se posee los recursos para utilizar algún *software* pagado como Predpol. Es por ello, que Carabineros de Chile y la Universidad de Chile mediante el CEAMOS en vista de la creciente tendencia en utilizar herramientas científicas para predecir riesgo criminal, está desarrollando e implementando una solución integral en las plataformas de Carabineros de Chile, con el objetivo de generar modelos predictivos dinámicos y proveer predicciones en cada turno policial. A raíz de lo anterior, se motiva este trabajo de tesis el cual aborda la caracterización de riesgo criminal mediante funciones espacio-temporales incluyendo elementos fenomenológicos espaciales.

1.2. Alcances

Esta tesis aborda la estimación del riesgo criminal considerando como fuente principal de datos, la ocurrencia de eventos criminales para un periodo de tiempo que contemplado entre 2011 y 2015 y para algunas comunas de la Región Metropolitana. La metodología para la estimación del riesgo criminal utiliza la información de eventos criminales históricos así como los ocurridos recientemente con el objetivo de estimar superficies que representan el riesgo criminal y realizar predicciones para diferentes ventanas de tiempo, también en una versión más refinada de la metodología se incluye información geográfica como la ubicación de ca-

lle para mejorar la estimación y predicción del riesgo. En esta tesis se presenta un modelo espacio-temporal del riesgo criminal con el cual puede realizarse predicciones dinámicas, sin embargo, el foco se encuentra en el estudio de los modelos espaciales que se desprenden de la metodología, donde se estudia en detalle la influencia de los diferentes tipos de parámetros espaciales y se proponen mecanismos para poder controlarlos en pos de mejorar el desempeño predictivo general. Esta metodología fue desarrollada en el marco de un proyecto con Carabineros de Chile y la Universidad de Chile mediante CEAMOS, por lo cual, por temas de confidencialidad (ver Sección correspondiente) solo se presenta un caso de estudio detallado para la comuna de Puente Alto y resultados generales para diferentes comunas de la Región Metropolitana, por lo que las validaciones exhaustivas no se presentan en esta tesis. Cabe mencionar que para el desarrollo de estos modelos, no se han cuestionado los datos proporcionados por Carabineros de Chile, vale decir, se ha asumido que la escritura de los datos no presentan errores significativos que puedan minar la capacidad de los modelos predictivos. Cabe mencionar que esta metodología se diseña con el propósito de ser flexible y escalable para cualquier zona de estudio siempre y cuando se tenga una cantidad adecuada de eventos. Por último no se aborda la distribución de las fuerzas policiales ni los recorridos que deban tomar a causa de las predicciones.

1.3. Objetivos

El objetivo principal de este trabajo de tesis consiste en el diseño y prueba de concepto¹ de un modelo que sea capaz de **caracterizar** el riesgo criminal espacio-temporal de una zona de estudio determinada con el fin de realizar predicciones del riesgo criminal para diferentes ventanas espaciales y temporales, haciendo uso de datos reales de comunas de la Región Metropolitana. Todo lo anterior enfocado a asistir la labor de las fuerzas policiales en asignar sus recursos para evitar nuevos focos de criminalidad. A partir del objetivo anterior se desprenden los siguientes objetivos específicos:

- Diseñar una metodología capaz de caracterizar el riesgo criminal espacio-temporal captando tendencias históricas (conocimiento *a priori*) y tendencias recientes (conocimiento *a posteriori*). Además debe contar con pasos definidos para realizar una clara caracterización espacial y temporal.
- Diseñar un modelo predictivo a partir de las caracterizaciones del riesgo criminal y con ello métricas adecuadas para medir desempeño predictivo.
- Desarrollar mecanismos de ajuste de parámetros que permitan el uso flexible de la metodología: En ese sentido la tesis aborda específicamente **cómo influyen los parámetros de la caracterización espacial del riesgo criminal condicionado a una caracterización temporal fija**² en la calidad de las predicciones que por supuesto se miden a través de métricas *ad hoc*.

¹Ver la sección de Acuerdos de Confidencialidad.

²Si bien en esta tesis solo se aborda la influencia de los parámetros espaciales en la calidad de las predicciones, debido a la naturaleza espacio-temporal del modelo, se hace necesario la mención de la metodología temporal, sin embargo, dado los alcances de esta tesis, se aborda en detalle el análisis espacial, dejando los análisis temporales respectivos en la tesis inédita del Sr. Pablo Fuentes [6].

- Determinar los alcances y limitaciones de la metodología y de qué manera puede ser perfeccionada.

1.4. Hipótesis de Trabajo

Dados los objetivos y alcances anteriores, se tienen las siguientes hipótesis de trabajo:

- El uso de la teoría de kernels para modelar y predecir la función de intensidad de riesgo espacio-temporal resulta idóneo para los Delitos de Mayor Connotación Social (DMCS), específicamente los agrupados de Robo con Fuerza y Robo con Violencia. En ese sentido, en el marco del proyecto relacionado con Carabineros de Chile, se acepta un modelo como idóneo si en términos predictivos obtiene al menos para un mes de validación el 35 % de aciertos promedio para un 10 % del área geográfica de mayor riesgo.
- Con respecto al trabajo central de tesis: Es posible determinar criterios y/o heurísticas sobre los parámetros de la caracterización espacial para mejorar el desempeño predictivo de la metodología completa. En ese sentido, se propone que una caracterización idónea del riesgo criminal se logra con una correcta elección de la cantidad de componentes espaciales, la cantidad de iteraciones del algoritmo de *Expectation Maximization*, la forma de las matrices de covarianza y las restricciones sobre la razón entre los valores propios de éstas para las GMM.
- Además, dependiendo de las características de las zonas de estudio, es posible obtener mejores desempeños para la metodología propuesta en su versión dinámica, y para otros casos la versión estática.
- Es posible mejorar la caracterización espacial de riesgo criminal y por ende el desempeño predictivo si se condiciona el dominio espacial a la geometría de las calles de la zona de estudio.
- Por último es posible realizar comparaciones con algoritmos de la literatura científica, pese a que sus análisis estén efectuados en otras zonas de estudios.

1.5. Acuerdos de Confidencialidad

Este trabajo de tesis se desprende de un proyecto de Carabineros de Chile con CEAMOS, que es un centro cuyo objetivo consiste en generar investigación en materia de criminología con impacto a nivel internacional. Debido a que este proyecto involucra procesamiento de datos reales, involucrando delitos de alta connotación social, se firmó un acuerdo de confidencialidad con el objetivo de resguardar información relacionada con eventos criminales registrados en las comunas de la Región Metropolitana desde 2011 hasta 2016. Además la confidencialidad se extiende para los algoritmos desarrollados y las validaciones en detalle realizados en las diferentes comunas, por lo cual en este documento solo se presenta un estudio detallado de Puente Alto y resultados generales para diferentes comunas. De esta manera, para efectos prácticos, el algoritmo presentado se cataloga como una prueba de concepto.

1.6. Estructura de la Tesis

- En el Capítulo 2 se presentan los métodos y técnicas actuales para la caracterización y predicción de riesgo criminal adicionando un marco teórico para abordar la metodología del Capítulo 3.
- En el Capítulo 3 se presenta la metodología incluyendo preprocesamiento de datos, procedimiento espacial, temporal y de predicción.
- En el capítulo 4 se presentan los resultados previos de la metodología para una comuna en particular más un estudio de los factores y parámetros más significativos de la caracterización espacial. Adicionalmente se generan resultados para diferentes comunas de la Región Metropolitana y se comparan desempeños. Finalmente se discuten los resultados obtenidos y se agrega una sección de comparación de resultados con otros algoritmos de la literatura científica.
- En el Capítulo 5 se concluye sobre los resultados y se discute las mejoras a la metodología desde el punto de vista espacial.

Capítulo 2

Estado del Arte y Marco Teórico

2.1. Metodologías para Estimación de Riesgo Criminal

Las fuerzas policiales con el objetivo de combatir el crimen de manera más eficaz han puesto a disposición tecnologías de georeferenciación de eventos criminales y características del entorno mediante plataformas tipo GIS (*Geographical Information Systems*), para facilitar la visualización y estudio de patrones delictuales (más detalles ver [20] y [21]). Con ello, el concepto de *Predictive Policing* como bien se aborda en el libro de W. Perry *et al.* [8] surge de manera espontánea, cuyo objetivo es la aplicación de técnicas cuantitativas para predecir eventos criminales y así asistir la labor policial. La naturaleza de estos métodos así como sus objetivos son variados como se desprende de la tabla de la Figura 2.1 (ver trabajo de W. Perry *et al.* [8]) la cual muestra distintas soluciones para los problemas planteados en la predicción de crímenes.

En la Figura 2.1 se mencionan los métodos más comúnmente utilizados para la estimación y predicción de riesgo criminal, no quitando que existan formulaciones con más variaciones. En ese sentido, como se irá constatando en esta tesis, el método planteado se encuentra en la categoría de *Spatiotemporal analysis methods*, pero contando con una cantidad de datos moderada. En esta metodología también se incorpora tanto el uso de datos históricos como recientes (para la identificación de *Hot-Spots*) y también permite el uso de características geográficas como calles. Lo que si es importante mencionar es que no se incluye el comportamiento de victimización repetida en espacio-tiempo (ítem *Near-repeat modeling*), siendo reemplazado por un análisis espectral ya sea en base a kernels periódicos o también mediante transformada Wavelet.

El *mapeo criminal*, el cual bien se analiza en el trabajo de S. Chainey *et al.* [9], consiste en la representación del riesgo criminal de manera espacial aglutinando ventanas temporales de ocurrencia de delitos. Este análisis es estático, puesto que no se modela el cambio del riesgo criminal explícitamente en el tiempo. Solo es posible dar cuenta de la variación de patrones espaciales en el tiempo realizando estos mapas para diferentes ventanas temporales. Este enfoque es útil para realizar estudios de largo plazo, sin embargo no aporta una clara fenomenología en la estimación y predicción de delitos en el corto plazo. En el trabajo de S.

Law Enforcement Use of Predictive Technologies: Predicting Crimes

Problem	Conventional Crime Analysis (low to moderate data demand and complexity)	Predictive Analytics (large data demand and high complexity)
Identify areas at increased risk		
Using historical crime data	Crime mapping (hot spot identification)	Advanced hot spot identification models; risk terrain analysis
Using a range of additional data (e.g., 911 call records, economics)	Basic regression models created in a spreadsheet program	Regression, classification, and clustering models
Accounting for increased risk from a recent crime	Assumption of increased risk in areas immediately surrounding a recent crime	Near-repeat modeling
Determine when areas will be most at risk of crime	Graphing/mapping the frequency of crimes in a given area by time/date (or specific events)	Spatiotemporal analysis methods
Identify geographic features that increase the risk of crime	Finding locations with the greatest frequency of crime incidents and drawing inferences	Risk terrain analysis

Figura 2.1: Métodos matemáticos para predecir ocurrencia de crímenes de acuerdo al problema que se desea solucionar. Fuente: Libro de W. Perry *et al.* [8].

Chainey *et al.* [9] además se desarrollan diferentes métodos para la generación de *Hot-Spots* espaciales como se aprecia en la Figura 2.2, destacando el método de KDE que constituye un método no paramétrico, donde se sitúa un kernel (generalmente Gaussiano) en cada dato considerando un ancho de banda determinado¹. Como la complejidad de los modelos espaciales determinados por medio de KDE dependen de la cantidad de datos, puede ser costosa la evaluación del riesgo y difícil la determinación de *Hot-Spots* de importancia, para lo cual se debe recurrir a métodos de *clustering*.

Siguiendo la línea de estimación mediante KDE, también se puede destacar el trabajo realizado en Chile por CONASET (Comisión Nacional de Seguridad de tránsito) llamado Análisis Espacio Temporal de los Siniestros de Tránsito en el Gran Santiago [23] en donde estudia la caracterización espacial de siniestros a lo largo de una base de datos desde el 2009 a 2013 en las comunas del Gran Santiago. Lo interesante de este trabajo es que se logra comparar año a año los patrones de siniestros de tránsito mediante correlaciones sobre los *Hot-Spots* para así establecer tendencias de éstos, y con ello permitir la elaboración de estrategias de largo plazo para evitar futuros siniestros. Este trabajo puede dar cuenta de la evolución de los *Hot-Spots* con un análisis estático año a año, sin embargo no se incluye un modelo predictivo con componentes temporales explícitas que permita predecir los siniestros futuros como sí se realiza en esta tesis.

¹También existen modelos donde se emplean KDE espacio-temporales como el trabajo de T. Nakaya *et al.* [22].

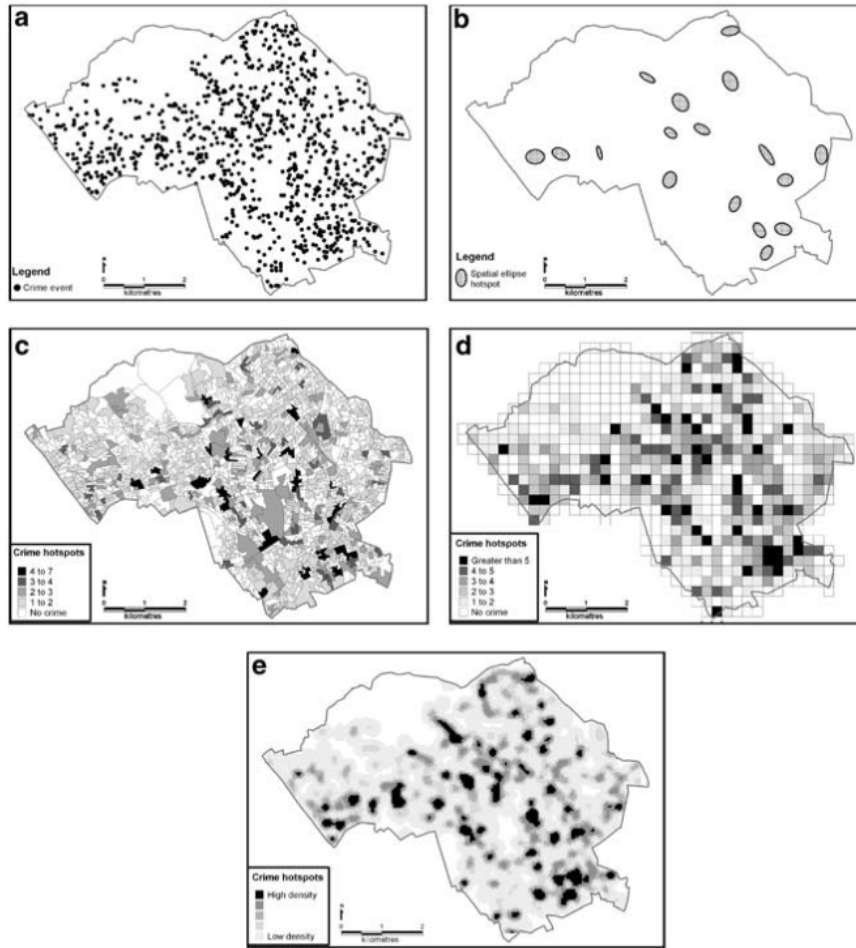


Figura 2.2: Diferentes métodos para la construcción de *Hot-Spots*. a) ocurrencia de eventos criminales, b) estimación de elipses mediante la desviación estándar de los delitos a su centro, c) designación de riesgo según zonas administrativas policiales, d) mapeo del territorio a través de grillas y finalmente en e) la estimación del riesgo mediante KDE (*Kernel Density Estimation*). Fuente: S. Chainey *et al.* [9].

En contraste con métodos no paramétricos como KDE surge la modelación mediante GMM (*Gaussian Mixture Models*) que constituyen modelos paramétricos en la estimación del riesgo, puesto que se tiene una cantidad finita de kernels (generalmente Gaussianas multivariadas) a ponderar y sumar, y por tanto la complejidad del entrenamiento se puede independizar de la cantidad de datos. De esta manera en implementaciones en línea resulta especialmente ventajoso el uso de GMM, puesto que dependiendo de la cantidad de recursos disponibles se fija la cantidad de parámetros a entrenar. En esta tesis se aborda la caracterización espacial de modelos utilizando GMM, introduciéndose métricas para estimar la cantidad de kernels óptimos pero además mecanismos para lograr captar adecuadamente *Hot-Spots* espaciales y que además permitan una correcta caracterización temporal.

En contraste con los métodos anteriores que no incluyen explícitamente la evolución temporal del riesgo criminal, se encuentran los que los que estudian los eventos criminales como procesos puntuales espacio-temporales (G. Mohler *et al.* [10]). En el trabajo de G. Mohler *et al.* [10] se equipara la modelación y predicción de la ocurrencia de delitos con la de even-

tos naturales como los terremotos, que si bien son fenómenos diferentes, pueden modelarse mediante una función de intensidad (o función gatillante), la cual posee dos componentes: la intensidad de fondo, que corresponde a la superficie en el espacio que se mantiene invariante en el tiempo y que en el caso de riesgo criminal puede interpretarse como el riesgo histórico de una zona de estudio; y los componentes auto-excitados, los cuales son fuertemente dependiente del tiempo y modelan el comportamiento posterior a la ocurrencia de un evento. Por lo general el componente auto-excitado modela la ocurrencia de eventos posteriores a la ocurrencia de un evento inicial, como en el caso de un terremoto y sus réplicas, o en riesgo criminal como un delito y las victimizaciones repetidas en tiempo y espacio. Este enfoque es completamente no paramétrico, donde se pueden determinar patrones espacio-temporales interesantes como se aprecia en la Figura 2.3 en donde se ha calculado las intensidades de fondo temporal y espacial para la zona de San Fernando Valley en Los Angeles, Estados Unidos, pero a expensas de un gran costo computacional, representando una desventaja si se considerara como un método en línea.

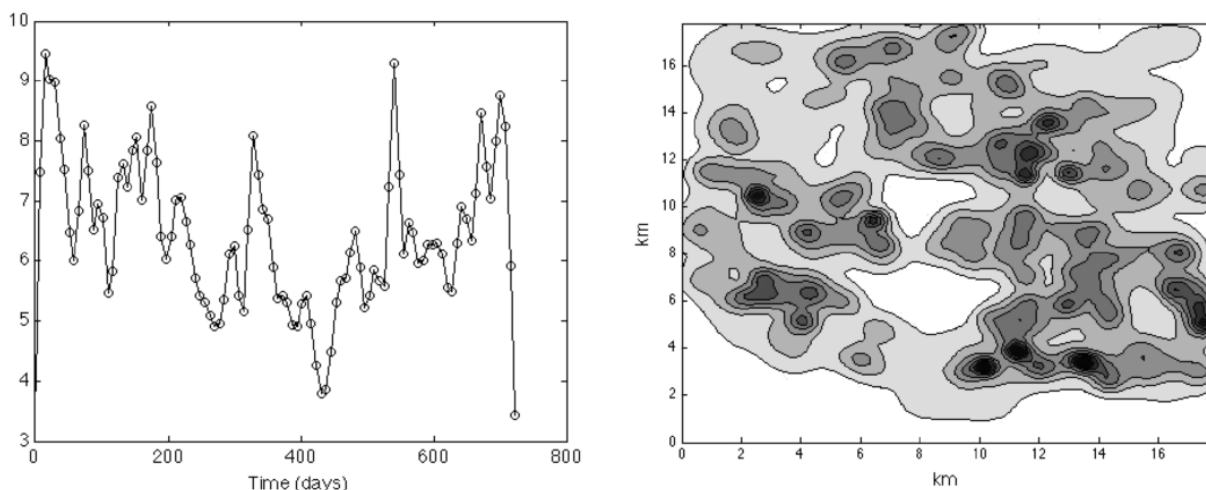


Figura 2.3: Estimación de las intensidades de fondo para el caso temporal y espacial. Fuente: [10].

La ecuación con la que estiman el riesgo es la siguiente:

$$\lambda(x, y, t) = \mu(x, y) + \sum_{\{k:t_k < t\}} g(t - t_k, x - x_k, y - y_k) \quad (2.1)$$

donde $\mu(x, y)$ es la intensidad de fondo y g es un kernel centrado en los puntos espacio-temporales t_k , x_k e y_k y la sumatoria representa las componentes auto-excitada que deben estimarse. Cabe mencionar que un enfoque similar llamado *Prospectivo* se pone a prueba en el trabajo de K. Bowers *et al.* [24] y que aprovecha de manera explícita que la ocurrencia de eventos criminales está ligada a victimizaciones repetidas en espacio y tiempo, y es uno de los algoritmos de comparación con el ya descrito en el trabajo de G. Mohler *et al.* [10]. A diferencia de los métodos anteriores, y como se había ya mencionado, en esta tesis se propone un análisis espacial paramétrico dado por GMM, y en términos temporales se estudia los componentes periódicos de las series de tiempo en cada *Hot-Spot* identificado.

La función de intensidad también puede modelarse mediante procesos Gaussianos como en el trabajo de Flaxman. S. [7]. En este trabajo, se asume que la ocurrencia de eventos

criminales sigue un proceso de Poisson con parámetro λ , el cual se estima como un proceso Gaussiano espacio-temporal como se describe en la siguiente ecuación:

$$y_{s,t} | \lambda(s, t) \sim \text{Poisson}(\exp(f(s, t)) \cdot e_s) \quad (2.2)$$

donde $y_{s,t}$ es la cantidad de delitos ocurridos en una ventana t y grilla s del espacio, $\lambda(s, t)$ es el parámetro del proceso el cual varía en cada t y s , y es equivalente a $\exp(f(s, t)) \cdot e_s$, donde $f(s, t)$ es un proceso Gaussiano y e_s es un ponderador que depende solamente de la grilla s . Los procesos Gaussianos se caracterizan por una función de media (generalmente en cero) y una de covarianza habiendo múltiples maneras de elegirlo. En este caso se tiene que:

$$f \sim \mathcal{G}(0, K) \quad (2.3)$$

donde K es la función de covarianza en donde se incluye las dependencias espacio-temporales por medio de kernels:

$$K((s, t), (s', t')) = k_s(s, s') + k_t(t, t') + k_{st}((s, t), (s', t')) + k_p(t, t') \quad (2.4)$$

donde:

- $k_s(s, s')$ es un kernel espacial para la grilla s con centro en s'
- $k_t(t, t')$ es una función exponencial radial
- $k_p(t, t')$ es un kernel periódico (ver en Sección 2.5.1)
- $k_{st}((s, t), (s', t')) = k_s(s, s') \cdot k_t(t, t')$

Lo interesante del trabajo de Flaxman. S. [7], es el uso de kernels espacio-temporales para la modelación de la función de covarianza. En esta tesis se recoge la idea de reconstrucción de señales espacio-temporales mediante kernels, en donde para el caso temporal al igual que el trabajo de Flaxman. S. [7] se utiliza kernels periódicos con el objetivo de encontrar los periodos en que la ocurrencia de crímenes tiene mayor injerencia. En secciones posteriores se explica con mayor detalle que no basta con solo encontrar los periodos de importancia, sino que además se debe encontrar el centro del kernel con su respectivo periodo y ancho de banda, haciendo más compleja la optimización.

Además de los enfoques anteriores, también existen modelos dinámicos que interpretan el riesgo criminal mediante partículas, como bien se desarrolla en la memoria de la Srta. P. Flores [25] y que fue inspiración inicial de este trabajo de tesis en desarrollar algoritmos de predicción dinámicos para comunas de la Región Metropolitana. A diferencia del método de esta tesis la caracterización del riesgo espacial en la memoria de la Srta. P. Flores [25] se basa en la discretización de GMM en partículas. Con ello la actualización dinámica de las partículas de riesgo criminal se plantea como el resultado de experimentar una fuerza de atracción Gaussiana por parte de nuevos eventos criminales conforme el tiempo avanza. Por su parte la predicción del movimiento de las partículas de riesgo se realiza por medio de un filtro RLS (*Recursive Least Squares*) considerando el movimiento histórico de éstas. En esa misma línea la predicción se mejora en P. Flores *et al* [26] donde el autor de esta tesis también participa, considerando ahora que los eventos criminales en términos temporales están enmarcados en un proceso de Poisson. Cabe mencionar que en los trabajos anteriores

consideran que la caracterización de riesgo criminal existe una influencia de largo plazo (dada por datos históricos de ocurrencia de delitos) y una de corto plazo (data por datos recientes de ocurrencia de delitos). Estos conceptos también se utilizan en esta tesis para general modelos *a priori* y *a posteriori* del riesgo criminal.

Por su parte, el uso de elementos geográficos para mejorar las predicciones de riesgo criminal también es utilizado actualmente, como el uso de servicios como se desprende de la memoria de la Srta. P. Flores [25] donde el supuesto principal es que ciertos tipos de delitos tienden a ocurrir cerca de servicios tales como hospitales, bancos, almacenes, etc.; pero también el uso de calles como se desarrolla en el trabajo de K. Kempinska *et al.*[27] y A. Okabe *et al.* [28]. En esta tesis también se aborda el uso de calles como condicionantes geométricos en la caracterización y predicción del riesgo criminal, pero se aborda de manera superficial y con el objetivo de demostrar que el método propuesto es flexible como para agregar este tipo de características.

En la misma línea anterior, hay modelos dinámicos que no solo incorporan ocurrencia de delitos sino que también explotan el uso de elementos demográficos y socio económicos en la estimación de riesgo criminal, como en el trabajo de X. Wang *et al.* [4] en donde se utiliza regresiones logísticas para estimar el riesgo criminal en una zona de estudio dividida en grillas. Para ello se utiliza el concepto de ST-GAM (*Spatio-Temporal Generalized Additive Model*) siendo más general que GMM. La ventaja de este tipo de modelos es la interpretabilidad ya que es posible detectar la influencia de las variables de entrada en la generación de riesgo criminal. La desventaja sin embargo, es que se debe hacer un preprocesamiento mayor de los datos en cada zona de estudio.

En resumen, esta tesis se enfoca en realizar caracterización de la función de intensidad de eventos criminales mediante modelos generados por GMM, donde a modo innovador **se introducen métricas para determinar la cantidad de parámetros a entrenar y mecanismos para controlar la forma de las matrices de covarianza**. Además en términos temporales se describe un método en base a kernels temporales y análisis espectral mediante Wavelet y por supuesto, se presenta el modelo predictivo con funciones espacio-temporales explícitas. Además en el modelo predictivo se considera una ponderación dada por datos históricos (generación de modelos *a priori*) y otra ponderación por datos recientes (generación de modelos *a posteriori*). También se analiza el efecto de condicionar el riesgo criminal a calles en las predicciones realizadas, y finalmente todos los análisis predictivos realizados se evalúan con métricas *ad hoc* que se explican en secciones posteriores, y que en artículos de la literatura también utilizan por lo cual hace posible realizar comparaciones. A continuación se realiza una pequeña introducción de procesos puntuales y el desarrollo de los conceptos necesarios para comprender la metodología desarrollada

2.2. Introducción a Procesos Puntuales

Muchos procesos físicos y sociológicos de interés toman la forma de eventos que ocurren de manera puntual en el espacio-tiempo, incluyendo eventos criminales [10], terremotos [29] e incluso fallas en estructuras [30]. Estos son denotados procesos puntuales espacio temporales

o *Spatio-Temporal Point Processes* (STPPs como lo indican sus siglas en inglés). A diferencia de procesos que pueden ser medidos en intervalos regulares y para espacios designados como el tráfico vehicular [31] y la temperatura de la superficie oceánica [32], los STPPs son solamente observados cada vez que se monitorea algún tipo de incidente. Debido a lo anterior, modelar STPPs constituye una tarea desafiante puesto que son dispersos, es decir, que la cantidad de eventos registrados en todo el dominio espacio-temporal resulta cercano a cero. Por ejemplo en el caso de eventos criminales, pueden haber intervalos importantes de tiempo entre delitos, que puede ser incluso de más de un día para una determinada zona. Es por esto que métodos estadísticos típicos no son los más adecuados para modelar STPPs dispersos o *sparse STPPs* (SSTPPs).

Afortunadamente, existen desarrollos matemáticos adecuados para describir SSTPPs [33], en donde los eventos puntuales son generados a partir de una función de intensidad continua, la cual no es observada directamente sino solo por el gatillamiento de eventos. En términos probabilísticos, esta función de intensidad puede interpretarse como una función de distribución generada a partir de un proceso estocástico continuo variante espacio-temporalmente. Generalmente los modelos basados en SSTPPs estiman la función de intensidad como la cantidad fluctuante de dónde los datos pueden generarse. Cabe mencionar que cuando los eventos puntuales tienen otros atributos o etiquetas se les denomina *marked STPPs*, por ejemplo al clasificar la ocurrencia de diferentes tipos de eventos criminales.

Un objetivo fundamental de los métodos predictivos es poder identificar las regiones en donde es más probable hallar una mayor cantidad de ocurrencia de eventos puntuales en el futuro, como por ejemplo predecir áreas en donde va a ocurrir una gran cantidad de eventos criminales. Tales regiones se denominan *Hot-Spots* o puntos calientes. Los *Hot-Spots* dichos son útiles en el quehacer de las fuerzas policiales ya que así es posible ubicar recursos de manera eficiente para evitar futuros eventos. Algunos ejemplos claves en predicción corresponden a la anticipación de la ocurrencia de terremotos [1], como en el estudio de riesgo criminal [10], [8]. Más detalles de procesos puntuales y un *framework* de análisis de modelos en [2].

2.3. Definición de Procesos Puntuales

En general los procesos puntuales pueden definirse como un conjunto dado por:

$$\mathcal{D} = \{(\nu_i, l_i)\}_{i=1}^N \quad (2.5)$$

Donde $\nu_i \in \mathbb{R}^D$ y $l_i \in \mathcal{L}$, con \mathcal{L} como el conjunto de las posibles etiquetas que caracterizan los elementos del proceso, de esta manera un STPP es el caso particular cuando:

$$\mathcal{D}' = \{(x_i, y_i, t_i, l_i)\}_{i=1}^N \quad (2.6)$$

Donde $x_i, y_i \in \mathbb{R}$, $t_i \in \mathbb{R}^+$ y $l_i \in \mathcal{L}$. Un proceso puntual puede ser descrito en términos probabilísticos considerándose como un proceso estocástico discreto, donde cada índice espacio-temporal que se esté dando corresponde a una variable aleatoria en particular, es decir, que $\nu_i = (x_i, y_i, t_i) \sim p_i(\nu) \forall i = 1, \dots, N$ existiendo la distribución conjunta $p_{1, \dots, N}(\nu_1, \dots, \nu_N)$.

Asimismo un proceso puntual puede corresponder al muestreo arbitrario de un proceso estocástico continuo $\nu(x, y, t)$. También un STPP puede describirse alternativamente por su *counting process* $N(x, y, t)$ que se define como [34]:

$$N(x, y, t) = |\{n : n \in [-\infty, x] \times [-\infty, y] \times [0, t]\}| \quad (2.7)$$

donde $|\cdot|$ es la cardinalidad del conjunto. Como $N(x, y, t)$ es función del proceso puntual, también es un proceso estocástico.

2.4. Función de Intensidad

Como se había mencionado anteriormente, un proceso puntual es gatillado por una función de intensidad $\lambda(x, y, t)$ que se puede definir como el límite de la esperanza de la razón de acumulación de eventos en una determinada vecindad espacio-temporal, es decir:

$$\lambda(x, y, t) = \lim_{\Delta t, \Delta x, \Delta y \rightarrow 0} \mathbb{E} \left\{ \frac{N(x + \Delta x, y + \Delta y, t + \Delta t) - N(x, y, t)}{\Delta t \Delta x \Delta y} \middle| \mathcal{D}' \right\} \quad (2.8)$$

donde $\mathbb{E}\{\cdot\}$ es la esperanza condicionada al STPP estudiado [33].

2.5. Estimación de Función de Intensidad

Antes de presentar métodos de estimación de funciones de intensidad mediante kernels como *Kernel Density Estimation*, *Expectation Maximization* y *Gaussian Mixture Models*, se define el concepto de **kernel** y se dan ejemplos de algunos de ellos.

2.5.1. Definición de Kernels y Tipos

Un kernel, también llamado función de covarianza, es un mapeo que tiene como argumento dos variables x y x' y como salida un valor real. Usualmente x y x' son elementos de un espacio Euclidiano, pero también pueden definirse sobre otros espacios siempre y cuando se cumplan ciertas propiedades. Muchas veces un kernel se interpreta como una métrica que compara dos valores o una medida de similitud. Formalmente se define como $k(x, x') \in \mathbb{R}$ para $x, x' \in \mathcal{X}$, típicamente como una función simétrica $k(x, x') = k(x', x)$ y no negativa $k(x, x') \geq 0$. Los kernels son utilizados en variados problemas de procesamiento de señales, ya sea para su reconstrucción o caracterización de patrones, clasificación y/o detección. A continuación se presentan los kernels más utilizados:

Kernel Gaussiano

$$k(x, x') = \exp\left(-\frac{1}{2}(x - x')^T \Sigma^{-1}(x - x')\right) \quad (2.9)$$

donde Σ es la matriz de covarianza definida positiva y que en caso de ser diagonal se tiene:

$$k(x, x') = \exp\left(-\frac{1}{2} \sum_{j=1}^D \frac{1}{\sigma_j^2} (x_j - x'_j)^2\right) \quad (2.10)$$

donde σ_j se interpreta como el ancho de banda característico para la dimensión j . Ahora si Σ es diagonal con valores iguales se genera un kernel isotrópico es decir:

$$k(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right) \quad (2.11)$$

donde σ es el ancho de banda. Más detalles en [35].

Cabe mencionar que el kernel Gaussiano también es definido positivo, ya que se puede representar mediante una serie de potencia formando parte del caso polinomial infinito dimensional, y de esta manera el vector ϕ también sería infinito dimensional.

Kernel Periódico

Usado para caracterizar patrones con periodicidades y modelos con mezcla de kernels y procesos Gaussianos [3] tiene la siguiente expresión:

$$k(x, x') = \sigma^2 \exp\left(-\frac{2}{l^2} \sin^2\left(\pi \frac{x - x'}{p}\right)\right) \quad (2.12)$$

En la Figura 2.4 se aprecian tres curvas para diferentes valores de l . En este caso x' se ha fijado como centro, p corresponde al periodo de los *peaks* y l actúa como el ancho de banda.

2.5.2. *Kernel Density Estimation*

Kernel Density Estimation o estimación de densidad por kernel es un método no paramétrico para estimar funciones de densidad para variables aleatorias a partir de muestras experimentales. Este método es el conversor por excelencia de procesos puntuales a funciones de intensidad. Desde un punto de vista estadístico este método se interpreta como una manera de construir histogramas pero además teniendo control de la función que suaviza esta construcción, es decir, que no se remite solo al uso de funciones uniformes (uso de barras) sino que a una amplia gama incluyendo funciones triangulares, Gaussianas e incluso

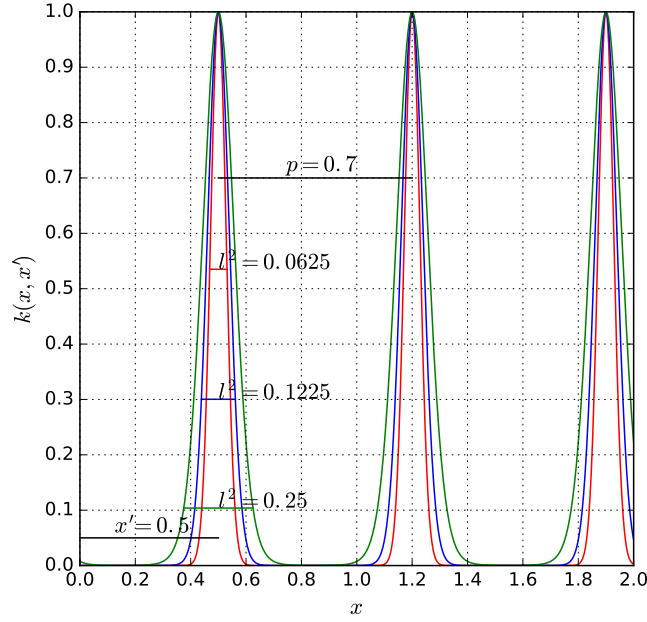


Figura 2.4: Kernel periódico con $x \in [0, 2]$, $x' = 0.5$, $\sigma = 1.0$, $p = 0.7$ y $l = [0.25, 0.35, 0.5]$

Epanechnikov que tienen en común ser kernels. Este procedimiento parte del supuesto que existen muestras i.i.d. (independientes e idénticamente distribuidas) $(x_1, \dots, x_n) \sim f$ con f desconocida, donde el estimador \hat{f} se calcula como:

$$\hat{f}_\theta = \frac{1}{n} \sum_{i=1}^n K(x, x_i | \theta) \quad (2.13)$$

donde K es un kernel con un vector de parámetros θ y centro x_i . Por lo general los kernels a trabajar son de la forma:

$$K(x, x_i | \theta) = K(x - x_i | h) = \frac{1}{h} K\left(\frac{x - x_i}{h}\right) \quad (2.14)$$

donde $x - x_i$ indica que es un kernel invariante al desplazamiento y que $h > 0$ es el ancho de banda del kernel. La elección del ancho de banda para estos kernels no es un tema menor, existiendo diferentes criterios para su cómputo como por ejemplo el MISE (*Mean Integrated Squared Error*), así como heurísticas como la regla de Silverman. Para el caso de estimación de funciones de intensidad en criminología, usualmente el ancho de banda se fija en cientos de metros para caracterizaciones espaciales (victimización cercana) y en horas para caracterizaciones temporales (victimización repetida).

2.5.3. Modelo por Mezcla de Gaussianas

Un Modelo por Mezcla de Gaussianas o en inglés *Gaussian Mixture Model* (GMM) es un caso particular de Modelo por Mezcla, el cual en estadística corresponde a generar distribuciones de probabilidad a partir de la suma ponderadas de otras. De esta manera, un GMM se construye a partir de la suma ponderada de distribuciones Gaussianas, que en el caso multidimensional cada Gaussiana posee un vector de media y una matriz de covarianza. Matemáticamente se define como:

$$\text{GMM}(x) = \sum_{i=1}^N \alpha_i \mathcal{N}(x|\mu_i, \Sigma_i) \quad (2.15)$$

donde:

$$\mathcal{N}(x|\mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^D |\Sigma|}} \exp\left(-\frac{1}{2}(\mu - x)^T \Sigma^{-1}(\mu - x)\right) \quad (2.16)$$

D es la dimensión, Σ es una matriz definida positiva y $|\Sigma|$ es su determinante, además:

$$\sum_{i=1}^N \alpha_i = 1 \quad (2.17)$$

Los GMM's son ampliamente utilizados para caracterizar datos multimodales, es por ello que en criminología los GMM's son muy adecuados para caracterizar concentraciones delictivas espaciales o *Hot-Spots*. Si bien tanto un GMM como el método de KDE pueden estimar funciones de intensidad a partir de un proceso puntual, a diferencia del método de KDE, los GMM's constituyen un modelo paramétrico, donde una vez establecida la cantidad de componentes a entrenar, solo basta estimar los parámetros de las Gaussianas (medias y matrices de covarianza) y los pesos α_i de cada una. Métodos estandarizados existen para su entrenamiento, como es el caso del algoritmo de *Expectation Maximization* (EM) que en Anexos 5.1.6 se explica en más detalle. Los parámetros de un GMM se pueden interpretar físicamente para el caso de caracterización espacial de riesgo criminal. Los centros se consideran *Hot-Spots* y dependiendo de la forma de la matriz de covarianza es la manera en cómo se distribuye el riesgo (ya sea esférico o en elipses con dirección arbitraria de los vectores propios en el plano).

2.5.4. Métricas para Análisis de Modelos Espaciales

A continuación se definen métricas para evaluar los modelos espaciales constituidos por GMM. Las métricas apuntan a dar un diagnóstico más visual de la forma de los kernels espaciales y su posicionamiento que solo evaluar la verosimilitud con el algoritmo EM. Esto, debido a que estos modelos tienden a quedarse en mínimos locales cuando se entrenan sus parámetros.

Silhouette

Corresponde a una métrica que mide cuán similar es un dato con respecto a su *cluster* (cohesión) comparada con otros *clusters* (separación). Su rango varía entre -1 y 1, en donde valores altos indica que el dato se corresponde con su cluster teniendo poca correspondencia con otros y viceversa. Si cada dato de un conjunto posee valores altos, entonces el *clustering* se considera adecuado. Para el caso de GMM, *silhouette* cuantifica cuán bien los componentes espaciales explican los eventos criminales analizados. Asumiendo que existen N clusters diferentes y datos con etiqueta $\{1, \dots, i, \dots, M\}$, entonces el *silhouette-score* del dato i se calcula como:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2.18)$$

donde $a(i)$ es la distancia promedio de i con respecto a todos los miembros del mismo *cluster*, mientras que $b(i)$ es la menor distancia promedio del dato i con respecto a *clusters* donde i no es miembro. Finalmente el *silhouette-score* SS es el promedio del *silhouette-score* de todos los datos [36]. En términos interpretativos, *silhouette-score* asigna un puntaje del grado de intersección de los *clusters* para los datos entrenados, donde 1 indica que los *clusters* no se intersecan y -1 indica intersección total de ellos. De esta manera el óptimo se alcanza cuando:

$$N_x = \operatorname{argmax}_{N \in \{N_{\min}, \dots, N_{\max}\}} SS(\text{GMM}_N, \mathcal{E}) \quad (2.19)$$

$$(2.20)$$

donde $SS(\text{GMM}_N, \mathcal{E})$ es el *score-silhouette* para una GMM de N componentes y datos \mathcal{E} y $\{N_{\min}^{pr}, \dots, N_{\max}^{pr}\}$ son los enteros a explorar en la optimización.

Para ilustrar esta métrica, en la Figura 2.5 se presenta un ejemplo donde se generan datos a partir de 4 clusters y mediante el entrenamiento de una GMM se fija como parámetro 10 componentes. Como se puede apreciar, el promedio del coeficiente de Silhouette es de 0.3 aproximadamente, habiendo datos con coeficiente de Silhouette negativos dejando en evidencia la poca representatividad de su propio cluster.

Asimismo en la Figura 2.6 se muestra el desempeño de esta métrica cambiando el número de componentes de entrenamiento, demostrado que en el caso de dos clusters se obtiene el óptimo pese a que originalmente se habían generado datos con 4. Esto ocurre debido que para esta métrica resulta óptimo aglutinar 3 clusters en uno solo (zona sur-oeste de la Figura 2.5). Luego el siguiente valor máximo corresponde a 4 que conforma un máximo local. En las siguientes secciones se muestra una optimización multi-objetivo para poder disminuir el problema de sub-estimación de cantidad de componentes en el entrenamiento de una GMM.

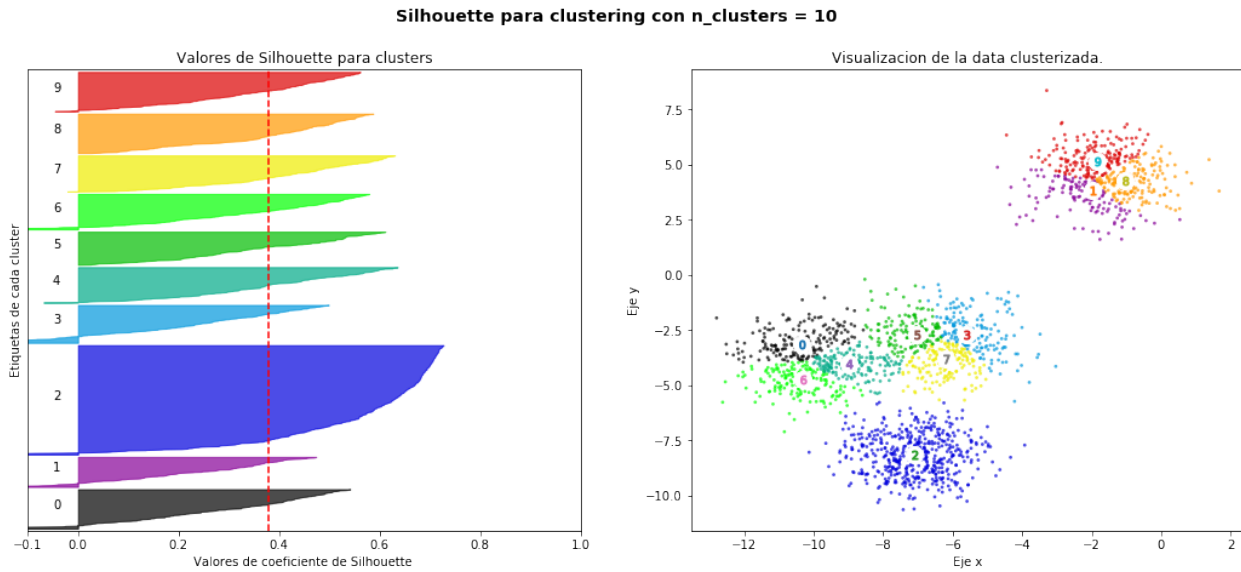


Figura 2.5: Coeficiente de Silhouette para cada dato de entrenamiento considerando datos artificiales generados por 4 clusters versus un clustering mediante 10 clusters.

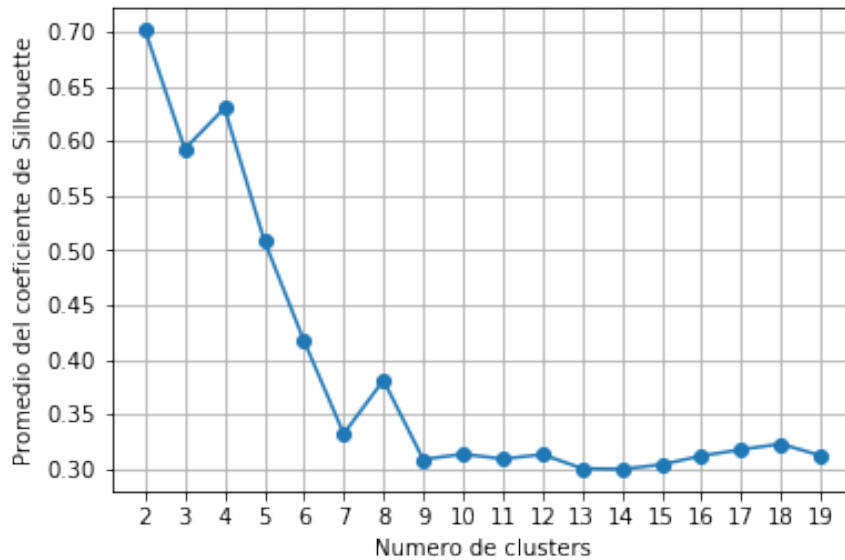


Figura 2.6: Promedio del coeficiente de Silhouette para diferente cantidad de clusters o componentes del algoritmo GMM.

Desviaciones

Dado el entrenamiento de un modelo GMM de N componentes con un conjunto de datos \mathcal{E} , se define los conjuntos de datos que se asignan a cada componente de la GMM como:

$$\mathcal{K}_c = \left\{ \nu \in \mathcal{E} : c = \underset{l \in \{0, \dots, N-1\}}{\operatorname{argmax}} R(s, l) \forall s \in \{0, \dots, |\mathcal{E}| - 1\} \right\} \forall c \in \{0, \dots, N - 1\} \quad (2.21)$$

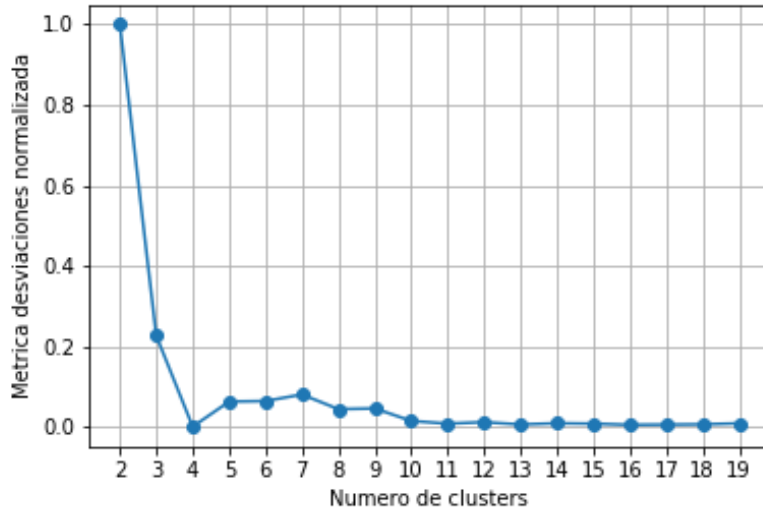


Figura 2.7: Métrica de desviaciones para diferente número de clusters o componentes del algoritmo GMM.

donde R definida como la *responsabilidad* indica la probabilidad de que cada observación s pertenezca a un determinado componente de la GMM. Este criterio calcula la varianza de la distribución empírica de las cantidades de eventos asignados a cada \mathcal{K}_c . Esto porque muchas veces para escoger la cantidad de componentes para el entrenamiento de una GMM se hace en función de la cantidad de eventos criminales por kernel deseados (primer momento), pero no se observa el comportamiento de la varianza, por lo que una elección puede no ser la adecuada si existe una gran desviación, teniendo muchos kernels con una cantidad de delitos muy por abajo del promedio esperado. De esta manera el objetivo es asegurar que la distribución tenga la menor varianza posible para que así la cantidad de datos por kernels sea lo más homogénea, con ello el óptimo se alcanza cuando:

$$N_x = \underset{N \in \{N_{\min}, \dots, N_{\max}\}}{\operatorname{argmin}} \quad \hat{\sigma}^2(\hat{p}(\mathcal{K}_1, \dots, \mathcal{K}_N)) \quad (2.22)$$

donde $\hat{\sigma}^2$ es el estimador de la varianza empírica y $\hat{p}(\mathcal{K}_1, \dots, \mathcal{K}_N)$ es la distribución empírica para \mathcal{K}_j . El problema anterior es equivalente a maximizar la **Cohesión** de la distribución empírica, calculada como el inverso de la varianza:

$$N_x = \underset{N \in \{N_{\min}, \dots, N_{\max}\}}{\operatorname{argmax}} \quad 1/\hat{\sigma}^2(\hat{p}(\mathcal{K}_1, \dots, \mathcal{K}_N)) \quad (2.23)$$

Para el mismo conjunto de datos utilizado en la Figura 2.5 se calcula la métrica de desviaciones como se aprecia en la Figura 2.7 donde el mínimo se alcanza para las 4 componentes originales. Cabe mencionar que se podría dar el caso que para una gran cantidad de componentes el criterio de desviaciones tienda a alcanzar el mínimo, por lo que como se ve en las siguientes secciones se establece un criterio multi-objetivo para solucionarlo.

Aciertos

Este criterio consiste en calcular la fracción de eventos criminales dentro de las zonas de mayor riesgo señaladas por la función de riesgo. En primer lugar se debe fijar el porcentaje del área total que se considere de importancia y luego se debe buscar el corte transversal de la superficie de riesgo tal que la integral desde infinito hasta el corte sea equivalente a esa área. Dado que las fuerzas policiales tienen ya una grilla espacial que discretiza la zona de estudio, basta con integrar la superficie de riesgo en cada cuadrícula y luego escoger las grillas de mayor riesgo que representen un porcentaje fijo de la zona de estudio. Si se define el conjunto de grillas de una zona de estudio determinada como \mathcal{G} , entonces el riesgo integrado de una grilla $p_i \in \mathcal{G}$ se define como:

$$r_i \triangleq r(p_i) = \int_{\mathbf{Int}(p_i)} R(x, y) dx dy \quad (2.24)$$

donde R es la función de riesgo que se integra en el interior de p_i (i.e. $\mathbf{Int}(p_i)$), con ello $\sum_i^{N_g} r_i = 1$, y la métrica FA (Fracción de Aciertos o Aciertos) se define como:

$$\mathbf{FA}(r \in [0, 1]) = \frac{\left| \left\{ \nu \in \mathcal{E} : \nu \in \bigcup_{\{i: r < r_i\}} \mathbf{Int}(p_i) \right\} \right|}{|\mathcal{E}|} \quad (2.25)$$

donde \mathcal{E} es el conjunto de datos espaciales utilizados. El objetivo de este criterio es maximizar la fracción de eventos criminales acertados para las grillas seleccionadas, por lo que se busca:

$$N_x^{\text{PF}} = \underset{N \in \{N_{\min}, \dots, N_{\max}\}}{\operatorname{argmax}} \mathbf{FA}(r) \quad (2.26)$$

para un r fijo determinado.

Optimización Multi-objetivo: Silhouette + Aciertos

Considerando el entrenamiento de GMM con diferente cantidad de componentes, *Silhouette* tiende a privilegiar caracterizaciones con menor cantidad de componentes para evitar que los kernels se traslapen entre sí. Por otro lado el criterio de Aciertos tiende a privilegiar caracterizaciones con una gran cantidad de componentes ya que así se puede tener una representación más fiel del riesgo generado por los eventos criminales, sin embargo una cantidad excesiva de componentes puede incurrir en un sobre entrenamiento de los modelos, y con ello pérdida de la capacidad de generalización. De esta manera, contraponer estos criterios resulta en una optimización multi-objetivo de la forma:

$$N_x = \underset{N \in \{N_{\min}, \dots, N_{\max}\}}{\operatorname{argmax}} \gamma_1 \cdot \text{SS}(\text{GMM}_N, \mathcal{E}) - \gamma_2 \cdot \mathbf{FA}(r) \quad (2.27)$$

donde γ_1 y γ_2 son parámetros positivos que deben sumar uno.

Optimización Multi-objetivo: Silhouette + Desviaciones

Nuevamente considerando el entrenamiento de GMM con diferente cantidad de componentes, el criterio de Desviaciones tiende a privilegiar caracterizaciones con mayor número de componentes porque aun cierto punto la cantidad de eventos por kernels tiende a estabilizarse. En este caso se debe tener especial cuidado en no tender a caracterizaciones con muy pocos delitos por kernels, es por ello que Desviaciones se contrapone con *Silhouette* en una optimización multi-objetivo de la forma:

$$N_x = \operatorname{argmax}_{N \in \{N_{\min}^{pr}, \dots, N_{\max}^{pr}\}} \gamma_1 \cdot \text{SS}(\text{GMM}_N, \mathcal{E}) + \gamma_2 \cdot 1/\hat{\sigma}^2(\hat{p}(\mathcal{K}_1, \dots, \mathcal{K}_N)) \quad (2.28)$$

Otras Métricas de Evaluación

Existen otras medidas como el criterio de Información de Akaike (AIC) el cual penaliza modelos que estén *sobre-parametrizados*, en pocas palabras, para dos modelos que tienen igual desempeño (error cuadrático medio de la validación por ejemplo), siempre se va a preferir el modelo que tenga menor complejidad o menor cantidad de parámetros.

2.5.5. Métricas de Validación Predictiva

Una predicción espacio-temporal corresponde a la generación de una superficie de riesgo $R(x, y, t_n) \in \mathbb{R}^+$ con $x, y \in \mathcal{A} \in \mathbb{R}^2$ para un tiempo t_n de predicción. La superficie, vista como una función de intensidad o también como una densidad de probabilidad, debiese en términos ideales ser aquella que al ser muestreada genere los eventos criminales que ocurren realmente. Existen diferentes métricas que ayudan a evaluar cuán cerca o lejos se encuentra la superficie de riesgo de ser la ideal en los términos ya expuestos. Entre los criterios más utilizados corresponde a calcular el porcentaje de delitos acertados por la predicción considerando un área fija de cobertura [24]. El porcentaje de aciertos varía en la medida que el área de cobertura cambia, por lo cual se debe escoger una determinada área. En ese sentido, el Índice de Precisión Predictiva o *Predictive Accuracy Index* (PAI) [9], normaliza el porcentaje de aciertos del modelo por el porcentaje de área cubierta, interpretándose como una medida de densidad de eventos criminales acertados por unidad de área pintada. El problema de las dos medidas anteriores es que no consideran múltiples escenarios, es decir, cómo es el desempeño de la predicción para diferentes niveles de cobertura. De esta manera, se trabajan con dos métricas llamadas HRP_θ (*High Risk Percentage*) y TIP_θ (*True Incident Percentage*) [4], las cuales corresponden al porcentaje de grillas pintadas superiores a un umbral $\theta \in [0, 1]$, y el porcentaje de delitos acertados dentro de las grillas pintadas. Estas métricas se definen de la siguiente manera:

$$\text{HRP}(\theta) = \frac{\left| \left\{ g \in \mathcal{G} : r_g \triangleq \int_{\text{int}(g)} R(x, y, t_n) dx dy > \theta \right\} \right|}{|\mathcal{G}|} \triangleq \frac{|\Omega_\theta|}{|\mathcal{G}|} \quad (2.29)$$

$$\text{TIP}(\theta) = \frac{|\{\nu \in \mathcal{E} \cap \text{Int}(\Omega_\theta)\}|}{|\mathcal{E}|} \quad (2.30)$$

donde $|\bullet|$ es la cardinalidad de un conjunto, g corresponde al polígono que denota una grilla del conjunto \mathcal{G} , r_g es el valor del riesgo integrado de la grilla g , θ es el umbral de riesgo, ν son los eventos criminales del espacio de todos los eventos \mathcal{E} intersecados con los puntos interiores de Ω_θ . Cabe mencionar que como θ es un umbral para el riesgo integrado, necesariamente el área de cobertura es función del umbral. De hecho es explícitamente:

$$a(\theta) = \sum_{g \in \Omega_\theta} \text{Área}(g) \quad (2.31)$$

$$= a_0 |\Omega_\theta| \quad (2.32)$$

donde a_0 es el área de cualquier grilla de \mathcal{G} (puesto que todas son de igual valor). De esta manera con un área de cobertura fija se puede encontrar el rango de θ que cumpla con ello (ya que $a(\theta)$ no es inyectiva).

Por medio de las curvas HRP y TIP se puede realizar un estudio similar a la de las curvas ROC (Receiver Operating Characteristics), donde se grafica en el eje de las abscisas el HRP y en el de las ordenadas el TIP, de esta manera se obtiene una curva de desempeño, donde se indica la cantidad de aciertos por porcentaje de área cubierta para diferentes niveles de riesgo integrado o área cubierta. Típicamente para medir desempeño de esta clase de curvas se utiliza el AUC (*Area Under Curve*), métrica que indica cómo se comportan las predicciones en promedio considerando todas las coberturas posibles. El AUC corresponde a la integral entre 0 y 1 de la curva TIP/HRP. Idealmente el AUC debiese tener valor 1, es decir, para cualquier área de cobertura se busca un cien por ciento de aciertos, el peor caso es cuando el AUC vale 0.5, lo que significa básicamente lanzar una moneda en cada grilla para predecir un evento criminal.

Capítulo 3

Metodología de Caracterización de Riesgo Criminal Incluyendo Elementos Fenomenológicos

En este capítulo se da a conocer la metodología aplicada para la caracterización espacio-temporal de riesgo criminal y su correspondiente predicción. Esta metodología se puede resumir en cuatro pasos fundamentales que se pueden visualizar en el esquema de la Figura 3.1. La primera etapa corresponde al procesamiento de los datos, en donde se extraen los eventos criminales de interés condicionado a una ventana espacio-temporal definida, la segunda etapa es el procesamiento espacial, en donde se caracterizan patrones espaciales de los eventos criminales, los cuales pueden ser patrones históricos (caracterización *a priori* de los datos espaciales) como recientes (caracterización *a posteriori* de los datos) mediante GMM; la tercera etapa corresponde al procesamiento temporal, en donde se caracterizan patrones temporales de los eventos criminales condicionados al *clustering* realizado en el procedimiento anterior, es decir, se conforman series de tiempo por cada agrupación local de datos espaciales hecha anteriormente. El fundamento de unir el procesamiento espacial y temporal de esta manera radica en que existe una victimización repetida en espacio y tiempo, es decir, que eventos que ocurren en un *cluster* espacial, suelen repetirse en los alrededores, y lo mismo ocurre a nivel temporal, siguiendo un comportamiento similar a la fenomenología de ocurrencia de terremotos y sus réplicas [10]. Cabe mencionar que en términos temporales se buscan patrones que capturen tendencias periódicas que puedan ser proyectadas en el futuro. Estas proyecciones conforman la última etapa correspondiente a la de predicción en donde confluye tanto el análisis espacial como temporal. Finalmente las hipótesis anteriores son validadas con las métricas ya descritas en la sección anterior.

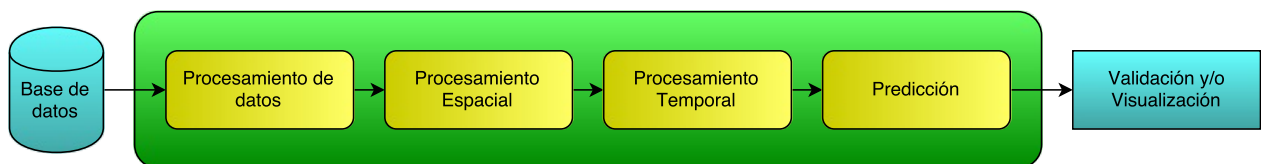


Figura 3.1: Metodología general para caracterización y predicción de riesgo criminal.

3.1. Procesamiento de Datos

Para la construcción de modelos predictivos es esencial tener claro los datos a utilizar para la etapa de entrenamiento. En este caso se consta de dos tipos de datos, el primero corresponde al set de eventos criminales etiquetados por tipo, fecha de ocurrencia, latitud y longitud, mientras que el segundo tipo corresponde a datos espaciales del ambiente como por ejemplo ubicación de servicios, calles, luminarias y paraderos. Desde un punto de vista frecuentista, los eventos criminales son adecuados en la caracterización de un modelo siempre y cuando se tenga una gran cantidad de ellos, debido a que vienen condicionados al entorno y accionar de las fuerzas policiales. Sin embargo, dependiendo del área de estudio, periodo de tiempo e incluso tipo de delito, la base de dato puede ser escasa para realizar un análisis basado solo en eventos criminales. De esta manera, los datos del entorno cobran una importancia similar a la de los eventos criminales, puesto que se puede establecer a partir del entorno cuál es el riesgo inminente de que ocurra un evento criminal. Cabe decir que este enfoque es Bayesiano puesto que se tiene un conocimiento previo o creencia del riesgo criminal el cual se va actualizando dependiendo de los eventos que vayan ocurriendo. A continuación se explica el tratamiento que se realiza a los eventos criminales sin considerar aún cómo tratar con los datos del entorno, puesto que estos últimos se utilizan para agregar mayor precisión al modelo.

3.1.1. Separación de Datos

En primer lugar, el set de datos se divide en tres grandes grupos que corresponde a la data *a priori*, data *a posteriori* y data de predicción o validación¹ La data *a priori* es aquella con la cual se realiza el entrenamiento grueso del algoritmo con el objetivo de captar tendencias y patrones espacio-temporales históricos, vale decir, las tendencias que han sido persistentes en el tiempo pese a los esfuerzos policiales en erradicarlas. En segundo lugar se encuentra la data *a posteriori*, que actúa como caracterizador de las últimas tendencias o patrones espacio-temporales de los eventos criminales. La data *a priori* por lo general tiene una extensión mínima de un año, y la *a posteriori* abarca periodos de dos a tres meses de datos [10] [2]. Por último la data de validación sirve para corroborar que las predicciones realizadas sean acordes al modelo entrenado. Dependiendo del enfoque predictivo varía el periodo de tiempo de validación, siendo para enfoques cortos de no más de un día, enfoques medianos de no más de una semana, y largos plazos que varían entre un mes e incluso un año.

En resumen para un conjunto de datos:

$$\mathcal{D} = \{(f_i, x_i, y_i) \in (\mathcal{F}, \mathbb{R}^2) : i = 1, \dots, N\} \quad (3.1)$$

¹Normalmente se trabaja con fechas y datos reales, por lo cual predecir a partir de una fecha actual real, no se cuenta con la validación para realizar las predicciones, por lo cual se deben establecer fechas actuales pasadas para validar los métodos.

en donde \mathcal{F} es el conjunto de todas la fechas factibles de la forma **YYYY-MM-dd HH:mm**, x es la longitud e y es la latitud, de esta manera se tienen los siguientes conjuntos resultantes:

$$\mathcal{D}_{\text{pr}} = \{\nu \in \mathcal{D} : f \in (f_{\text{ini}}, \dots, f_{\text{cut}}] \subseteq \mathcal{F}\} \quad (3.2)$$

$$\mathcal{D}_{\text{po}} = \{\nu \in \mathcal{D} : f \in (f_{\text{cut}}, \dots, f_{\text{act}}] \subseteq \mathcal{F}\} \quad (3.3)$$

$$\mathcal{D}_{\text{val}} = \{\nu \in \mathcal{D} : f \in [f_{\text{pred1}}, \dots, f_{\text{pred2}}] \subseteq \mathcal{F}\} \quad (3.4)$$

donde f_{ini} es la fecha inicial de la base de datos, f_{cut} es la fecha que separa el conjunto *a priori* del *a posteriori*, y $[f_{\text{pred1}}, \dots, f_{\text{pred2}}]$ es el intervalo de predicción cumpliéndose $f_{\text{act}} < f_{\text{pred1}}$.

3.2. Procesamiento Espacial

En esta metodología el procesamiento espacial puede realizarse de varias maneras, sin embargo se destacan dos enfoques principales como se aprecia en la Figura 3.2 que corresponden a la generación de un modelo GMM de los datos espaciales y la realización proyecciones geométricas. Este último enfoque consiste en rescatar elementos espaciales relevantes de la zona de estudio y proyectar los eventos criminales sobre estos elementos con el objetivo de condicionar el riesgo criminal espacial a éstos. Ejemplos de elementos geométricos son calles, servicios, luminarias, propiedad privada, etc. La idea es establecer una función de riesgo espacial que caracterice los elementos geométricos de tal manera que el riesgo sea uniforme dentro del polígono que describa el elemento, pero entre elementos el riesgo sea proporcional a la cantidad de eventos criminales proyectados. Si bien este enfoque tiene la ventaja de establecer las posibles causas de eventos criminales mediante elementos geométricos, puede resultar ser muy costoso computacionalmente por la variedad de elementos geométricos trabajados para la cantidad de eventos criminales existentes (que puede ser muy acotada). Es por ello, que el enfoque mediante GMM se suele usar para este tipo de análisis ya que no es costoso computacionalmente y por otro lado el entrenamiento de sus parámetros es muy eficiente mediante el algoritmo EM. La desventaja es que se debe tener una serie de consideraciones al entrenar GMM para que los resultados tengan una coherencia con la fenomenología del riesgo criminal. Estas consideraciones son tres: la primera consiste en establecer cuáles son las condiciones iniciales de los parámetros (cantidad de componentes, vectores de medias, matrices de covarianzas y pesos a utilizar), en segundo lugar establecer restricciones geométricas del problema como por ejemplo el tipo de matriz de covarianza a utilizar y restricciones sobre valores propios de la misma, y tercero los criterios de convergencia, en donde se debe tener en cuenta las condiciones de término del algoritmo para que entrene en un tiempo aceptable.

3.2.1. Elección Previa de Parámetros

El procesamiento espacial se realiza con las coordenadas geográficas de \mathcal{D}_{pr} y \mathcal{D}_{po} , vale decir sobre los conjuntos:

$$\mathcal{E}_{\text{pr}} = \{(x, y) \in \mathbb{R}^2 : (f, x, y) \in \mathcal{D}_{\text{pr}}\} \quad (3.5)$$

$$\mathcal{E}_{\text{po}} = \{(x, y) \in \mathbb{R}^2 : (f, x, y) \in \mathcal{D}_{\text{po}}\} \quad (3.6)$$

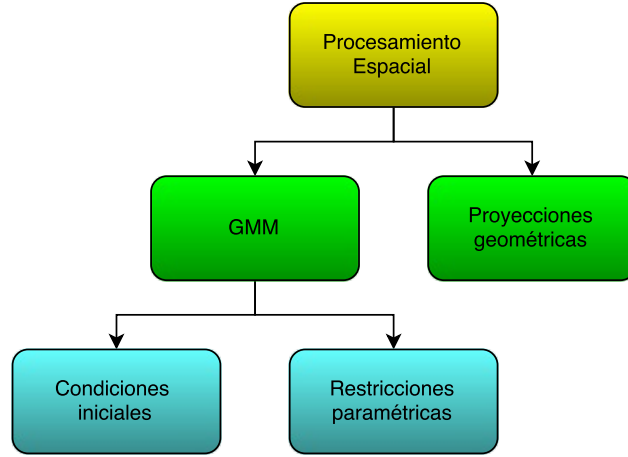


Figura 3.2: Esquema general para procesamiento espacial.

A partir de los conjuntos anteriores se modela la función de intensidad espacial mediante GMM, donde los parámetros son entrenados utilizando el algoritmo de EM descrito en el capítulo anterior. De esta forma se obtiene:

$$R_{\mathbf{x}}^{\text{pr}}(x) = \sum_{i=1}^{N_x^{\text{pr}}} \alpha_i^{\text{pr}} k_{\mathbf{x}}(x, \mu_i^{\text{pr}}, \Sigma_i^{\text{pr}}) \quad (3.7)$$

$$R_{\mathbf{x}}^{\text{po}}(x) = \sum_{i=1}^{N_x^{\text{po}}} \alpha_i^{\text{po}} k_{\mathbf{x}}(x, \mu_i^{\text{po}}, \Sigma_i^{\text{po}}) \quad (3.8)$$

$$(3.9)$$

donde:

$$k_{\mathbf{x}}(x, \mu, \Sigma) = \exp\left(-\frac{1}{2}(\mu - x)^T \Sigma^{-1}(\mu - x)\right) \quad (3.10)$$

y :

$$\Theta_{\mathbf{x}}^{\text{pr}} = \{(\alpha_i^{\text{pr}}, \Sigma_i^{\text{pr}}, \mu_i^{\text{pr}})\}_{i=1}^{N_x^{\text{pr}}} \quad (3.11)$$

$$\Theta_{\mathbf{x}}^{\text{po}} = \{(\alpha_i^{\text{po}}, \Sigma_i^{\text{po}}, \mu_i^{\text{po}})\}_{i=1}^{N_x^{\text{po}}} \quad (3.12)$$

Elección de Componentes Iniciales

Hasta ahora se asume que N_x^{pr} y N_x^{po} son conocidos, sin embargo son dos parámetros muy importantes que deben encontrarse además de los parámetros dados por $\Theta_{\mathbf{x}}^{\text{pr}}$ y $\Theta_{\mathbf{x}}^{\text{po}}$. Para tener una idea más clara de cómo influyen los distintos valores de estos dos parámetros basta con plantearse dos casos extremos. El primero consistiría en caracterizar el modelo espacial por un solo kernel, lo que conduciría a una caracterización general del riesgo criminal espacial sin tomar en cuenta patrones finos. Por el contrario, el segundo consistiría en colocar un kernel en cada evento criminal, lo que generaría un modelo con poca capacidad de generalización y por ende sin capacidad de caracterizar *Hot-Spots*. De esta manera, cabe preguntarse

cuál es el criterio para definir optimalidad sobre la cantidad de kernels espaciales que deben considerarse en los dos modelos descritos. La respuesta está estrechamente ligada a cuál es la cantidad de eventos criminales asociados por kernel espacial resulta ser adecuada para que su entrenamiento temporal sea satisfactorio: tener una baja densidad de eventos por kernel generaría un entrenamiento pobre de las series de tiempo, mientras que una alta densidad podría conducir a un entrenamiento complejo como para detectar patrones temporales periódicos bien definidos. En términos prácticos suele escogerse la cantidad de componentes como un porcentaje de los datos de entrenamiento (como métrica para cálculos *online*), sin embargo para dar más sustento a este tipo de elecciones se utiliza las optimizaciones multi-objetivos presentadas en el marco teórico como métricas de análisis *offline* como se presenta en la sección de resultados.

Elección de Medias

Por lo general el entrenamiento de una GMM se realiza para el conjunto de medias, covarianzas y pesos de las normales multivariadas, sin embargo, dependiendo del tipo de datos que se requiera ajustar, es posible reducir el espacio paramétrico Θ_{GMM} si se posee información de interés. En este caso, considerando que ciertos tipos de eventos criminales se caracterizan por ocurrir en segmentos de calles, intersecciones, zonas residenciales y servicios, se puede condicionar las medias al espacio generado por estos elementos espaciales. De esta manera se define $\mathcal{S}_1, \dots, \mathcal{S}_r$ como los conjuntos de centros para distintos tipos de elementos geométricos (por ejemplo \mathcal{S}_1 puede corresponder al conjunto de farmacias, \mathcal{S}_2 hospitales, \mathcal{S}_3 intersecciones de calles, etc.). Si se considera que un delito con etiqueta l_j tiene asociado $\bigcup_{k \in I_j} S_k$ tipos de centros donde I_j es el conjunto de índices asociado al delito l_j , entonces la cantidad de centros totales se calcula como:

$$N_{S|l_j} = \left| \bigcup_{k \in I_j} S_k \right| = \sum_{I_j} |S_k| \quad (3.13)$$

muchas veces ocurre que la cantidad de centros $N_{S|l_j}$ es mucho mayor a la cantidad de delitos l_j , por lo cual, se debe encontrar cuáles son los centros de mayor importancia sujeto a los delitos ocurridos. De esta manera, se debe realizar proyecciones de los delitos tipo l_j sobre $\bigcup_{k \in I_j} S_k$, en donde la proyección puede ser diferenciando el tipo de servicio, es decir, que el mismo evento debe ser proyectado a diferentes S_k , o sino una no diferenciada, en donde el delito se asigna únicamente a un centro independiente de su tipo. Cabe mencionar que la proyección se realiza por un criterio de mínima distancia. Posteriormente se escogen los centros que recibieron mayor cantidad de delitos escogiendo una cantidad fija mediante la optimización multi-objetivo ya explicada.

3.2.2. Restricciones Paramétricas

Elección de Matrices de Covarianza

Debido a que se está entrenando modelos en base a Gaussianas multivariadas, es posible elegir el tipo de matriz de covarianza de cada componente, que mediante un corte transversal de la superficie de riesgo del componente puede originar diferentes cónicas. Las cónicas a formar son circunferencias y elipses, donde los vectores propios de éstas últimas pueden adquirir cualquier dirección (ortogonales entre sí). Cabe mencionar que dependiendo del tipo de delito, algunas elecciones son más adecuadas que otras. Cuando los delitos son causados principalmente por elementos geométricos de tipo servicio, entonces las covarianzas debiesen ser diagonales para que adquieran una forma circular, lo que significaría una expansión de riesgo tipo radial. Si ahora los delitos ocurren en avenidas o calles, entonces la mejor manera de representar el riesgo es mediante componentes elípticas. De esta manera, identificar las calles más riesgosas correspondería a situar kernels Gaussianos en ellas, encontrando el ángulo de inclinación de éstas para así determinar los valores propios adecuados de las matrices de covarianza en cuestión.

Restricción Sobre Valores Propios

Muchas veces la caracterización por medio de GMM no resulta acorde a la fenomenología del riesgo criminal dada una cierta zona de estudio. Por ejemplo, a veces los componentes espaciales correlacionan zonas muy lejanas entre sí, o las elipses producidas no tienen una forma acorde a los elementos geométricos de la zona de estudio. Cuando se intenta caracterizar calles o avenidas, se busca que las componentes elípticas no estén desproporcionadas, es decir, que la razón de sus valores propios esté dentro de un umbral. Para explicar mejor esto, se considera la región de confianza de una Gaussiana multivariada de N dimensiones como:

$$(x - \mu)^T \Sigma^{-1} (x - \mu) \leq Q_{\chi_N^2}(p) \quad (3.14)$$

donde $Q_{\chi_N^2}(p)$ es el cuantil de la distribución *chi-squared* con N grados de libertad. Esta definición permite determinar la región elíptica que encierra p probabilidad, cuyos ejes se definen como:

$$\text{eje}_i(p|\Sigma) = \sqrt{\lambda_i Q_{\chi_N^2}(p)} \quad \forall i = 1, \dots, N \quad (3.15)$$

donde λ_i es el valor propio i -ésimo de la matriz de covarianza Σ . De esta manera, para un caso bidimensional, se puede establecer una restricción de desigualdad sobre los ejes de la forma:

$$r^2 \geq \frac{\text{máx}(\lambda_1, \lambda_2)}{\text{mín}(\lambda_1, \lambda_2)} \quad (3.16)$$

donde r es un parámetro de diseño. Hasta ahora solo se ha introducido una restricción, sin embargo ningún mecanismo para poder modificar las componentes espaciales que no cumplan con dicha restricción. A continuación se explica un mecanismo para solucionar el problema planteado.

Mecanismo para Cumplir con Restricción sobre Valores Propios

Considerando que el algoritmo EM ha entrenado los parámetros de una GMM para un cierto análisis espacial y que existe un conjunto de componentes espaciales que no cumplen con la restricción, dígame:

$$\mathcal{N}_{\mathcal{R}} = \left\{ (\alpha, \mu, \Sigma) \in \Theta_{\text{GMM}} : r^2 > \frac{\text{máx}(\lambda_1, \lambda_2)}{\text{mín}(\lambda_1, \lambda_2)} \right\} \quad (3.17)$$

el mecanismo consiste en generar dos componentes espaciales aledaños además del central (una a la izquierda y otra a la derecha de manera simétrica) a cada componente que no cumple con la restricción, donde cada componente aledaño está a $\gamma \cdot \text{máx}(\text{eje}_i(p|\Sigma))$ ($\gamma \in (0, 1]$) de μ es decir:

$$\mu_1 = \mu + \gamma \cdot \text{máx}(\text{eje}_i(p|\Sigma)) \cdot \text{máx}(\hat{v}_i) \quad (3.18)$$

$$\mu_2 = \mu - \gamma \cdot \text{máx}(\text{eje}_i(p|\Sigma)) \cdot \text{máx}(\hat{v}_i) \quad (3.19)$$

donde \hat{v}_i es el vector propio normalizado i -ésimo de la matriz de covarianza Σ . Por otro lado se generan nuevas matrices de covarianza Σ_1 y Σ_2 que se construyen de tal manera que sus ejes para una probabilidad p encerrada son:

$$\text{máx eje}_i(p|\Sigma_j) = \beta \cdot \text{máx}(\text{eje}_i(p|\Sigma)) \quad , \beta \in (0, 1] \quad , \forall j = \{1, 2\} \quad (3.20)$$

$$\text{mín eje}_i(p|\Sigma_j) = \text{mín}(\text{eje}_i(p|\Sigma)) \quad , \forall j = \{1, 2\} \quad (3.21)$$

Por último la matriz de covarianza (central) Σ se redefine como $\Sigma = \Sigma_j$. Este procedimiento se realiza para $\mathcal{N}_{\mathcal{R}}$ escogiendo β de tal manera que las nuevas matrices de covarianza cumplan con la restricción. En la Figura 3.3 se muestra el ejemplo para un componente espacial en donde $\frac{\lambda_1}{\lambda_2} \geq r^2$ (Figura 3.3a), por lo que se generan tres nuevas componentes de tal manera que β y γ cumplan con la restricción sobre los valores propios (Figura 3.3b).

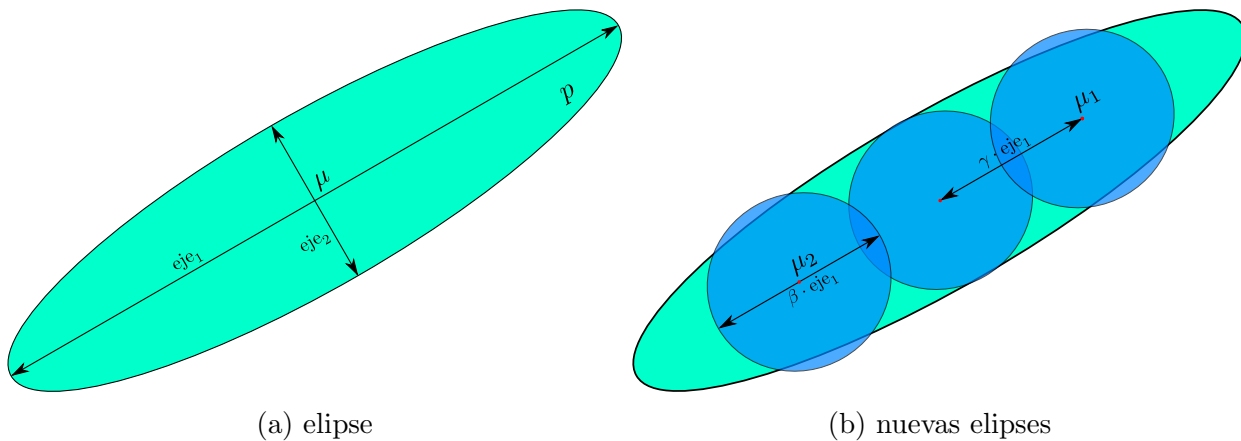


Figura 3.3: Ejemplo de mecanismo para cumplir con restricción sobre valores propios sobre un componente espacial Gaussiano.

3.3. Procesamiento Temporal

Si bien en este trabajo de tesis el procesamiento temporal no es el foco principal del desarrollo², de todas maneras debe generarse un procesamiento temporal con el fin de obtener resultados predictivos para validar la metodología. En la Figura 3.4 se aprecia el esquema general para realizar el procesamiento temporal que consta de dos procedimientos alternativos.

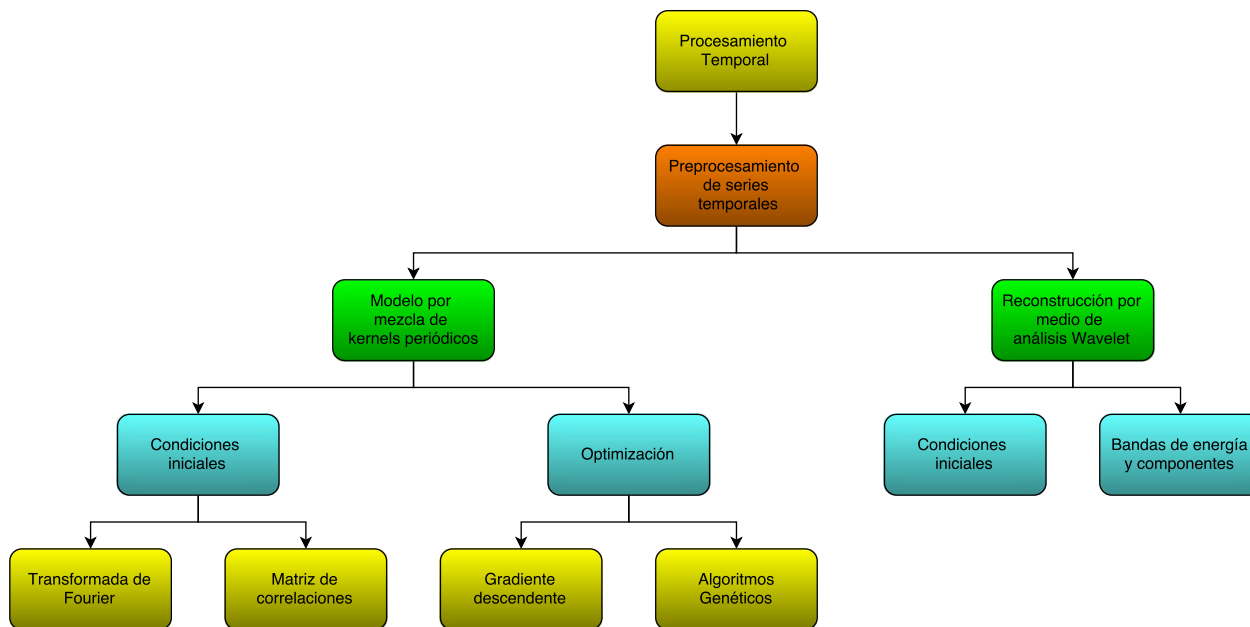


Figura 3.4: Esquema general para procesamiento temporal.

En primer lugar se construyen los conjuntos $\mathcal{K}_j^{\text{PF}}$ y $\mathcal{K}_k^{\text{Po}}$, los cuales corresponden a los eventos asignados a cada kernel del modelo *a priori* y *a posteriori* respectivamente (ver (2.21)). A

²puesto que los resultados se enfocan en hallar las relaciones entre parámetros espaciales en pos de mejorar resultados predictivos

partir de cada conjunto se construye una serie de tiempo, la cual debe ser procesada y finalmente utilizada para que pueda entrenarse un modelo temporal. Existe una enorme cantidad de técnicas para realizar predicción de series de tiempo. En este caso, basado en la naturaleza semi-periódica de los eventos criminales, se describen dos metodologías principales, la primera es la generación de modelos *a priori* y *a posteriori* por mezcla de *kernels* temporales periódicos y el segundo es la realización de un análisis de multi-resolución temporal mediante transformada Wavelet de las series de tiempo. Estos dos enfoques son comparados en detalle en la tesis del Sr. Pablo Fuentes [6] mostrando que el estudio mediante transformada Wavelet es el que menos costo computacional presenta versus el entrenamiento con mezcla de *kernels* periódicos, pero éste último presenta mejores resultados predictivos (considerando una caracterización espacial fija).

La generación de modelos por medio de mezcla de *kernels* periódicos, al igual que en el caso espacial, consiste en encontrar los parámetros de los *kernels* asociados de tal manera que pueda minimizarse alguna métrica, ya sea utilizando el argumento de Máxima Verosimilitud (ML) o minimización del Error Cuadrático Medio (MSE). Para realizar entrenamiento adecuado de los modelos temporales, existe una serie de consideraciones en cuanto a las condiciones iniciales para el algoritmo de optimización, en donde, dado que se trabajan con señales con periodicidades incorporadas, es posible realizar un análisis espectral por medio de la transformada de Fourier como también analizar frecuencias claves mediante el concepto de correlación. Los algoritmos de optimización basados en una métrica determinada pueden ser mediante un gradiente descendente (siempre y cuando se tengan buenas condiciones iniciales para evitar mínimos locales), o también algoritmo genéticos para tener claro el espacio paramétrico en el cual se está trabajando. De acuerdo a los dos criterios anteriores entonces se puede definir un término del algoritmo para entrenar los modelos temporales.

Por último, en cuanto a la representación de series de tiempo mediante análisis con transformada Wavelet, su fundamento radica en que los patrones temporales criminales exhiben un comportamiento periódico para diferentes escalas de tiempo (días, semanas, meses, etc), por lo cual, es posible representar una serie de tiempo solo utilizando las escalas temporales de mayor importancia, y descartar aquellas de menor energía que pueden interpretarse como ruido.

3.3.1. Preprocesamiento de Series Temporales

Ajustar Resolución Temporal

Los eventos criminales son STPP que ocurren en una determinada región espacio-temporal sin un muestreo regular en particular. En términos temporales, son series de tiempo que tienen una ocurrencia dispersa, es decir, que dada una resolución temporal determinada, existen muchos instantes en donde no hay presencia de delitos y muy pocas donde sí hay. Esto conlleva a que cualquier entrenamiento realizado para ajustar un modelo minimizando el MSE no sea adecuado [2]. Esto porque al estar *desbalanceada* la base de datos, la mayoría de los instantes la serie de tiempo sería nula, y las ocurrencias de delitos serían consideradas *outliers*. Cabe mencionar que este problema puede solucionarse disminuyendo la resolución

temporal, lo que ocasionaría un aglutinamiento de datos en pocos instantes de tiempo y por tanto una pérdida de precisión temporal al momento de realizar predicciones.

Suavización de las STPP

Otra forma de solucionar el problema sin cambiar la resolución temporal ya establecida, consiste en suavizar la serie de tiempo trabajada mediante el método de KDE. Este método se ocupa por lo general para estimar densidades de probabilidad. En este caso, se estaría estimando la función de intensidad de riesgo centrando kernels en cada evento. De esta manera si los conjuntos de interés son $\mathcal{K}_j^{\text{pr}}$ y $\mathcal{K}_k^{\text{po}}$ (ver (2.21)), entonces la versión suavizada de los procesos sería:

$$\bar{R}_j^{\text{pr}}(t) = \frac{1}{|\mathcal{K}_j^{\text{pr}}|} \sum_{t_0 \in \mathcal{K}_j^{\text{pr}}} K\left(\frac{t-t_0}{h}\right) \quad (3.22)$$

$$\bar{R}_j^{\text{po}}(t) = \frac{1}{|\mathcal{K}_j^{\text{po}}|} \sum_{t_0 \in \mathcal{K}_j^{\text{po}}} K\left(\frac{t-t_0}{h}\right) \quad (3.23)$$

donde h es el ancho de banda del *kernel* suavizador y t_0 es el centro de cada uno. Cabe mencionar que $K(\cdot)$ en este caso es Gaussiano, debido a su simplicidad de implementación y además que es coherente con la fenomenología del riesgo criminal, puesto que considera un riesgo previo ascendente antes de la ocurrencia del evento, luego riesgo máximo en la ocurrencia, y finalmente un riesgo descendente posterior.

Elección de Ancho de Banda

Este es un parámetro que debe elegirse cuidadosamente puesto que su tamaño determina con cuánta anticipación el riesgo criminal asciende antes de que ocurra el evento y también indica cuánto perdurará su intensidad una vez ocurra. Considerando un kernel exponencial de la forma:

$$k(t \geq 0, h \geq 0) = \exp\left(-\frac{t^2}{2h}\right) \quad (3.24)$$

se puede calcular el instante t^* para que el valor del kernel decaiga a $k^* \in (0, 1)$ como:

$$t^*(h) = \sqrt{-2h \ln k^*} \geq 0 \quad (3.25)$$

Considerando que $[0, t^*(h)]$ es un intervalo donde el riesgo criminal es significativo, necesariamente se busca que al momento de muestrear las series de tiempo se logre evaluar la mayor cantidad de puntos dentro de ese intervalo. De esta manera dado un ancho de banda fijo, la frecuencia mínima de muestreo de la serie de tiempo dada por el teorema del muestreo de Shannon debe ser:

$$f_{\min} = \frac{2}{t^*(h)} \quad (3.26)$$

Finalmente si se toma f_{\min} como parámetro de diseño, entonces $h = t^{*-1}(2/f_{\min})$.

3.3.2. Modelo por Mezcla de Kernels Periódicos

Una vez se obtienen las series de tiempo suavizadas $\bar{R}_j^{\text{pr}}(t)$ y $\bar{R}_j^{\text{po}}(t)$ se deben entrenar modelos por mezcla de kernels periódicos tanto *a priori* como *a posteriori*:

$$R_j^{\text{pr}}(t) = \sum_{i=1}^{N_t^{\text{pr}}} \delta_{ij}^{\text{pr}} k_{\mathbf{t}}(t, t_{ij}^{\text{pr}}, l_{ij}^{\text{pr}}, p_{ij}^{\text{pr}}) \quad (3.27)$$

$$R_j^{\text{po}}(t) = \sum_{i=1}^{N_t^{\text{po}}} \delta_{ij}^{\text{po}} k_{\mathbf{t}}(t, t_{ij}^{\text{po}}, l_{ij}^{\text{po}}, p_{ij}^{\text{po}}) \quad (3.28)$$

En este caso, los parámetros:

$$\Theta_{R_j^{\text{pr}}} = \{(\delta_{ij}^{\text{pr}}, t_{ij}^{\text{pr}}, l_{ij}^{\text{pr}}, p_{ij}^{\text{pr}})\}_{i=1}^{N_t^{\text{pr}}} \quad (3.29)$$

$$\Theta_{R_j^{\text{po}}} = \{(\delta_{ij}^{\text{po}}, t_{ij}^{\text{po}}, l_{ij}^{\text{po}}, p_{ij}^{\text{po}})\}_{i=1}^{N_t^{\text{po}}} \quad (3.30)$$

deben ser encontrados por medio de un proceso de optimización ya sea por gradiente descendente, en donde para agilizar cálculos se debe obtener el gradiente del modelo con respecto a los parámetros y además hallar estrategias para evitar mínimos locales. Por otro lado, se puede utilizar algoritmos genéticos como PSO, que si bien son robustos ante mínimos locales, son computacionalmente costosos para esta aplicación. Es por esto que se utilizan algoritmos en base a gradientes con estrategias para hallar buenas condiciones iniciales. Finalmente al igual que el caso del entrenamiento espacial, se debe obtener los parámetros N_t^{pr} y N_t^{po} , cuyo cálculo es explicado en las siguientes secciones.

Optimización

En este caso se buscan los parámetros que minimizan el MSE entre $R_j^{\text{pr}}(t)$ y $\bar{R}_j^{\text{pr}}(t)$ para $\mathcal{K}_j^{\text{pr}}$ como:

$$\Theta_j = \underset{\Theta_j}{\operatorname{argmin}} \sum_{t \in \mathcal{K}_j^{\text{pr}}} (R_j^{\text{pr}}(t) - \bar{R}_j^{\text{pr}}(t))^2 \quad (3.31)$$

$$= \underset{\Theta_j}{\operatorname{argmin}} \sum_{t \in \mathcal{K}_j^{\text{pr}}} \left(\sum_{i=1}^{N_t^{\text{pr}}} \delta_{ij}^{\text{pr}} k_{\mathbf{t}}(t, t_{ij}^{\text{pr}}, l_{ij}^{\text{pr}}, p_{ij}^{\text{pr}}) - \frac{1}{|\mathcal{K}_j^{\text{pr}}|} \sum_{t_0 \in \mathcal{K}_j^{\text{pr}}} K\left(\frac{t-t_0}{h}\right) \right)^2 \quad (3.32)$$

este problema de optimización se resuelve mediante un algoritmo de gradiente descendente calculando la derivada del MSE con respecto a los parámetros.

Condiciones Iniciales

Debido a que este problema de optimización consta de mínimos locales, es necesario realizar un análisis de las señales entrenadas para así encontrar las condiciones iniciales más adecuadas. En este caso se tiene como alternativa encontrar por medio de la transformada de Fourier de la serie de tiempo $\mathcal{K}_j^{\text{pr}}$ (lo mismo para el caso *a posteriori*) las componentes de frecuencia (o periodos) de mayor energía, dígame $\mathcal{P}_{\text{Fourier}}^{\text{pr}}$, sin embargo no queda claro qué centros t_{ij}^{pr} de los kernels temporales se estarían asignando a los periodos encontrados de $\mathcal{P}_{\text{Fourier}}^{\text{pr}}$, es por esto que surge el **método de correlaciones**. En donde para un ancho de banda fijo, y para una grilla de componentes de frecuencias y centros, se calcula la correlación de la serie de tiempo estudiada con respecto a un kernel periódico con parámetros fijos. Con lo anterior se puede determinar los pares de centros y componentes de frecuencias de mayor correlación y escogerlos como condiciones iniciales fijas del algoritmo de optimización. De esta manera se define:

$$\Omega_{\text{corrj}}^{\text{pr}} = \left\{ (t', p', c') \in (T, P, \mathbb{R}) : c' = \sum_{t'} k_{\text{t}}(t, t', l, p') \cdot \bar{R}_j^{\text{pr}}(t) \right\} \quad (3.33)$$

como el conjunto de parámetros de kernels periódicos (centros, periodos, correlación), donde P es un conjunto finito de componentes de periódicas y T es un conjunto finito de centros. Finalmente mediante el conjunto anterior se puede escoger los N_t^{pr} pares de componentes de mayor correlación:

$$\mathcal{PC}_{\text{corrj}}^{\text{pr}} = \left\{ (t', p') : (t', p', c') \in \underset{(N_t^{\text{pr}}, c')}{\text{max-sort}}(\Omega_{\text{corrj}}^{\text{pr}}) \right\} \quad (3.34)$$

donde max-sort es el operador que toma los N_t^{pr} elementos de mayor correlación de $\Omega_{\text{corrj}}^{\text{pr}}$.

como se aprecia en la Figura 3.5 se muestra visualmente la determinación de los periodos y centros de mayor correlación con la señal suavizada.

Finalmente con aquellos elementos se realiza la optimización anteriormente planteada en donde considerando un ancho de banda fijo de los kernels, solo hay que calcular las constantes (δ_{ij}) que multiplican los kernels periódicos.

3.3.3. Reconstrucción de Señales Mediante Wavelet

El objetivo de realizar una reconstrucción mediante Wavelet para señales de tiempo suavizadas $\bar{R}_j^{\text{pr}}(t)$ y $\bar{R}_j^{\text{po}}(t)$ consiste en determinar cuáles son las escalas de tiempo de mayor energía o importancia que se desprenden de éstas. Particularmente en el caso de riesgo criminal, se toma como hipótesis que la ocurrencia de delitos es cuasi-periódica, lo que significa

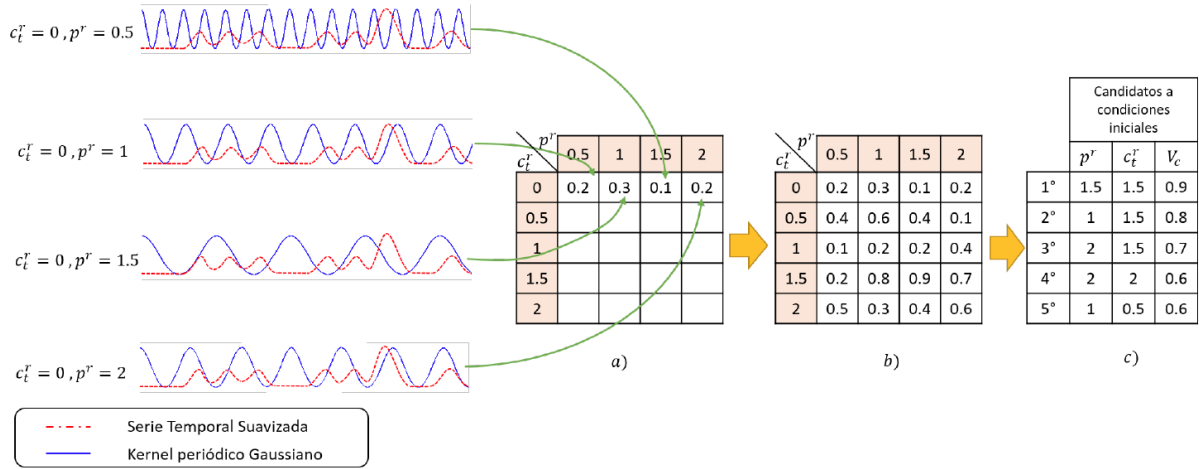


Figura 3.5: Método de correlaciones: para un conjunto de centros \mathbf{c} y periodos \mathbf{p} finitos se calcula la correlación entre la serie temporal suavizada $\bar{R}(t)$ con el kernel periódico para un determinado centro y periodo (construcción del conjunto $\Omega_{\text{corrj}}^{\text{pr}}$). Finalmente se ordenan según su grado de correlación y se escogen los elementos más correlacionados (construcción del conjunto $\mathcal{PC}_{\text{corrj}}^{\text{pr}}$). Fuente: tesis del Sr. Pablo Fuentes [6].

que si bien van a existir eventos fuera de los periodos de importancia, los patrones periódicos van a predominar los que a su vez se van a proyectar en el tiempo para realizar predicciones.

Basado en el estudio de C. Torrence *et al.* [11] y K.M. Lau *et al* [37] se estudian las series temporales *a priori* y *a posteriori* que se desprende de cada kernel espacial mediante la Transformada Wavelet Continua (TWC). La TWC de una señal discreta $x(k)$ se define como la convolución de ésta con una función base ψ :

$$W_n(s) = \sum_{n'=0}^{N-1} x(n')\psi^* \left(\frac{(n' - n)\delta t}{s} \right) \quad (3.35)$$

donde $*$ es el conjugado de la función base, n es el desplazamiento y s es la escala Wavelet. Las funciones base pueden adoptar diferentes formas, pero las más típicas son las que se aprecian en la Figura 3.6. Cabe mencionar que las funciones bases deben normalizarse tal que su energía sea unitaria para que la comparación entre escalas pueda hacerse directamente.

Debido a que las funciones bases son por lo general funciones complejas, la transformada $W_n(s)$ también lo es, por lo que se puede definir el Espectro de Energía Wavelet (EEW) como $|W_n(s)|^2$ como herramienta para analizar las escalas con mayor energía. Como se aprecia en la Figura 3.7 se escoge una serie temporal arbitraria asociada de un determinado kernel espacial del proceso *a priori* en donde se ha calculado el EEW para periodos de tiempo de 0 a 120 días³ y el histograma de la derecha, el cual tiene un ancho de banda fijo donde se ha calculado el aporte de las filas del EEW promedio. Analizando las bandas con mayor energía, se destaca el periodo 4.5 seguido de 7.5, 13.6 y 10.5 como los más importantes.

³Estos corresponden a los periodos de Fourier equivalentes a las escalas trabajadas

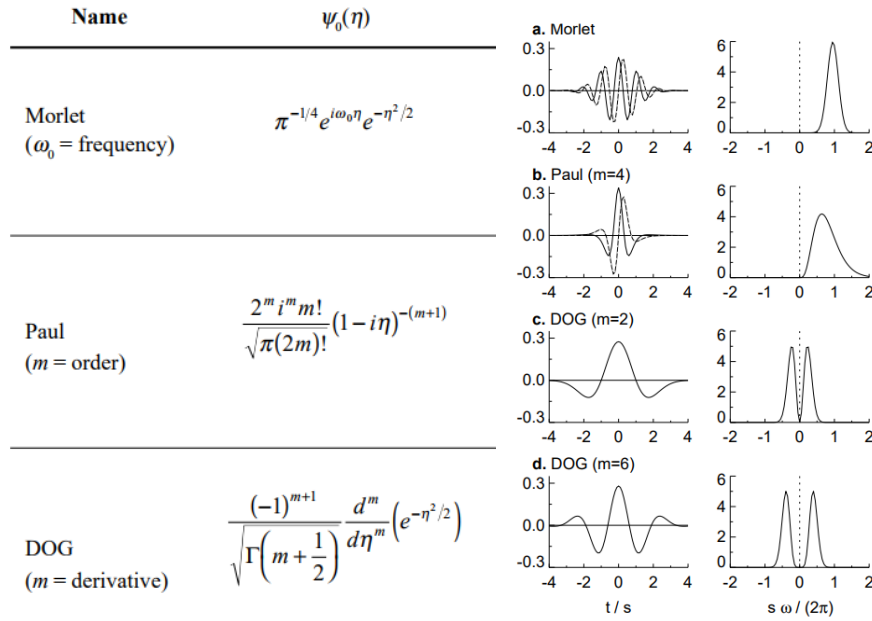


Figura 3.6: Ecuaciones (izquierda) que definen funciones bases típicas de Transformada Wavelet Continua con su respectiva forma (derecha). Fuente: Trabajo de C. Torrence *et al.* [11]

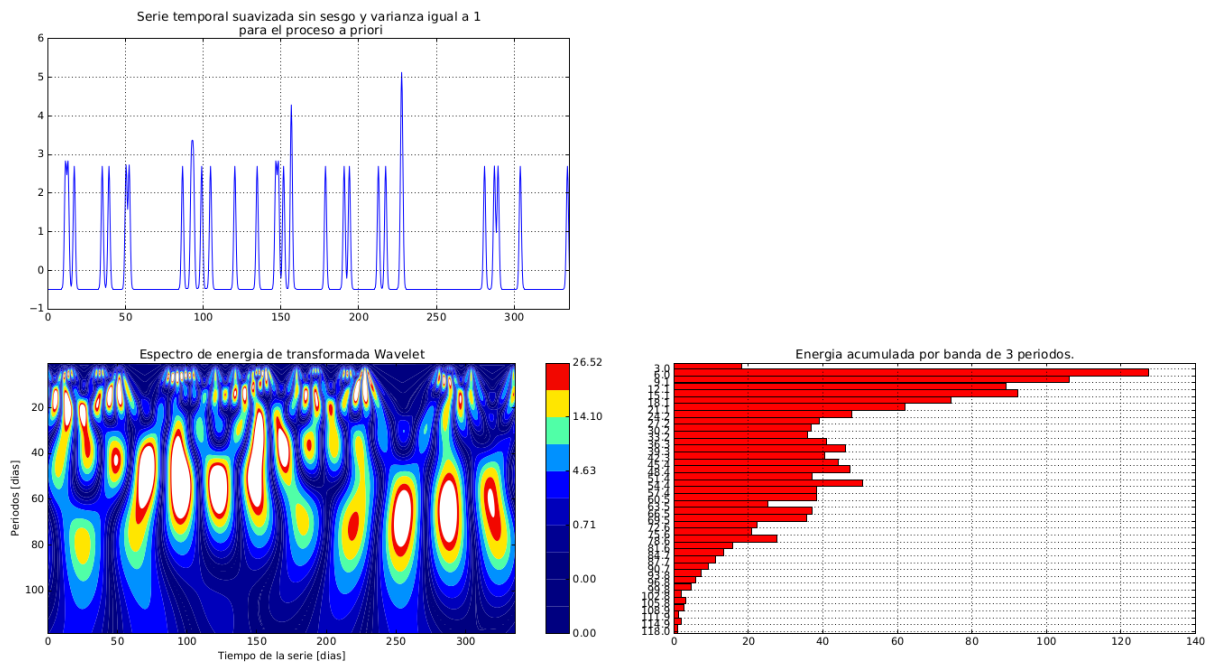


Figura 3.7: Señal temporal típica suavizada asociada a un kernel espacial con su respectivo EEW para periodos de tiempo de 0 a 120 días, y su histograma de energía promedio para bandas de ancho de 3 días. Fuente: tesis del Sr. Pablo Fuentes [6].

Seleccionando los 10 componentes más importantes se reconstruye la nueva señal mediante la **Antitransformada Wavelet**.

3.4. Procedimiento de predicción

El modelo general de predicción utiliza la caracterización espacio-temporal realizada en las secciones anteriores. Este modelo predictivo es una suma ponderada de la combinación de kernels espaciales y temporales como se expresa a continuación:

$$\begin{aligned}
R^{\text{pred}}(x, t) &= \lambda^{\text{pr}} \sum_{j=1}^{N_x^{\text{pr}}} R_j^{\text{pr}}(t) \alpha_j^{\text{pr}} k_{\mathbf{x}}(x, \mu_j^{\text{pr}}, \Sigma_j^{\text{pr}}) + \lambda^{\text{po}} \sum_{j=1}^{N_x^{\text{po}}} R_j^{\text{po}}(t) \alpha_j^{\text{po}} k_{\mathbf{x}}(x, \mu_j^{\text{po}}, \Sigma_j^{\text{po}}) \quad (3.36) \\
&= \lambda^{\text{pr}} \sum_{j=1}^{N_x^{\text{pr}}} \sum_{i=1}^{N_t^{\text{pr}}} \alpha_j^{\text{pr}} \delta_{ij}^{\text{pr}} k_{\mathbf{x}}(x, \mu_j^{\text{pr}}, \Sigma_j^{\text{pr}}) k_t(t, t_{ij}^{\text{pr}}, l_{ij}^{\text{pr}}, p_{ij}^{\text{pr}}) \\
&\quad + \lambda^{\text{po}} \sum_{j=1}^{N_x^{\text{po}}} \sum_{i=1}^{N_t^{\text{po}}} \alpha_j^{\text{po}} \delta_{ij}^{\text{po}} k_{\mathbf{x}}(x, \mu_j^{\text{po}}, \Sigma_j^{\text{po}}) k_t(t, t_{ij}^{\text{po}}, l_{ij}^{\text{po}}, p_{ij}^{\text{po}}) \quad (3.37)
\end{aligned}$$

donde $x \in \mathcal{A}$ correspondiente a la zona de estudio y $t \in [t_{\text{pred1}}, t_{\text{pred2}}]$ correspondiente al intervalo de predicción, además se destaca lo siguiente:

- λ^{pr} y λ^{po} : ponderación inter-procedimiento *a priori* y *a posteriori*.
- α_j^{pr} y α_j^{po} : ponderación espacial del j -ésimo kernel *a priori* y *a posteriori*.
- δ_{ij}^{pr} y δ_{ij}^{po} : ponderación para el j -ésimo kernel espacial con la i -ésima componente temporal *a priori* y *a posteriori*.

los parámetros inter-procedimiento *a priori* y *a posteriori* suelen escogerse como $\lambda \in (0, 1)$ y $1 - \lambda$ respectivamente donde λ se define como:

$$\lambda = \frac{|\mathcal{D}_{\text{pr}}|}{|\mathcal{D}_{\text{po}}|} \quad (3.38)$$

considerando que siempre $|\mathcal{D}_{\text{pr}}| > |\mathcal{D}_{\text{po}}|$. El modelo predictivo anterior se interpreta como un modelo por mezcla de kernels espacio-temporales de la forma:

$$k_{\mathbf{xt}}(x, t) = k_t(t, t', l, p) k_{\mathbf{x}}(x, \mu, \Sigma) \quad (3.39)$$

cabe mencionar que en la literatura existen más tipos de kernels espacio-temporales, en donde incluso es imposible separar la parte temporal de la espacial. En este caso, este modelo asume que un proceso pondera a otro y que además pueden entrenarse separadamente. Este modelo podría entrenarse con un solo algoritmo pero sería muy costoso computacionalmente poder implementarlo.

3.4.1. Proyección de Señales Pretéritas

Hasta ahora, las predicciones temporales se realizan simplemente evaluando las series de tiempo $R_j^{\text{pr}}(t)$ y $R_j^{\text{po}}(t)$ en un determinado $t \in [t_{\text{pred1}}, t_{\text{pred2}}]$, sin embargo, este procedimiento se puede sustituir reemplazando $R_j^{\text{pr}}(t)$ y $R_j^{\text{po}}(t)$ directamente por $\bar{R}_j^{\text{pr}}(\tau_1)$ y $\bar{R}_j^{\text{po}}(\tau_2)$. Con ello, se ahorra el entrenamiento de los modelos temporales, pero dado que $\tau_1 \in [t_{\text{ini}}, t_{\text{cut}}]$ y $\tau_2 \in [t_{\text{cut}}, t_{\text{act}}]$ no hay una manera natural de realizar predicción. Para ello, bajo el argumento de periodicidad, se realiza un *epoch folding* con factor de olvido de las series de tiempo estudiadas, es decir, considerando fases t_1 y t_2 y periodos p_1 y p_2 las series $\bar{R}_j^{\text{pr}}(\tau_1)$ y $\bar{R}_j^{\text{po}}(\tau_1)$ se particionan según:

$$\mathcal{P}_{\bar{R}_j^{\text{pr}}(\tau_1)} = \{ \bar{R}_j^{\text{pr}}(\tau_{11} \in [t_1, t_1 + p_1]), \dots, \bar{R}_j^{\text{pr}}(\tau_{1n} \in [t_1 + np_1, t_1 + (n-1)p_1]) \} \quad (3.40)$$

$$\mathcal{P}_{\bar{R}_j^{\text{po}}(\tau_2)} = \{ \bar{R}_j^{\text{po}}(\tau_{21} \in [t_2, t_2 + p_2]), \dots, \bar{R}_j^{\text{po}}(\tau_{2n} \in [t_2 + np_2, t_2 + (n-1)p_2]) \} \quad (3.41)$$

Luego las series de tiempo de predicción se construyen como:

$$R_j^{\text{pr}}(t|\gamma^{\text{pr}}) = \sum_{i=1}^n (\gamma^{\text{pr}})^i \bar{R}_j^{\text{pr}}(t - t_1 + i \cdot p_1) \quad (3.42)$$

$$R_j^{\text{po}}(t|\gamma^{\text{po}}) = \sum_{i=1}^n (\gamma^{\text{po}})^i \bar{R}_j^{\text{po}}(t - t_2 + i \cdot p_2) \quad (3.43)$$

t_1 se escoge como t_{pred1} y p_1 se escoge con una periodicidad de largo plazo como por ejemplo un año. Para el caso de t_2 se escoge como t_{pred2} y p_2 se escoge con una periodicidad de corto plazo, por ejemplo de una o dos semanas, incluso hasta un mes.

Cabe mencionar que estas nuevas series de tiempo pueden contener una cantidad de ruido importante, por lo cual, en vez de utilizar \bar{R}_j^{pr} y \bar{R}_j^{po} directamente, se puede utilizar la reconstrucción de estas señales mediante análisis de reconstrucción de señales mediante transformada Wavelet, que de hecho es el procedimiento utilizado en [6], donde la predicción de las señales temporales *a priori* se realiza al proyectar los valores asociados a un año antes de la fechas de predicción, mientras que las señales *a posteriori* son segmentadas por semanas considerando un intervalo de dos meses y ponderadas mediante un factor de olvido igual a 0.9.

3.5. Resumen del Algoritmo a Utilizar

En la Figura 3.8 se especifican los pasos del algoritmo para generar el modelo de predicción anteriormente mostrado. Cabe mencionar que dentro del procesamiento espacial, la obtención de medias y el mecanismo de restricción sobre covarianza no son pasos obligatorios; en la versión más simple del algoritmo basta con entregar los N_x^{pr} y N_x^{po} y entrenar los modelos mediante EM. En cuanto al procesamiento temporal, el preprocesamiento y suavización mediante KDE se debe hacer siempre sin importar si se utiliza el procedimiento mediante matriz de correlaciones o análisis wavelet.

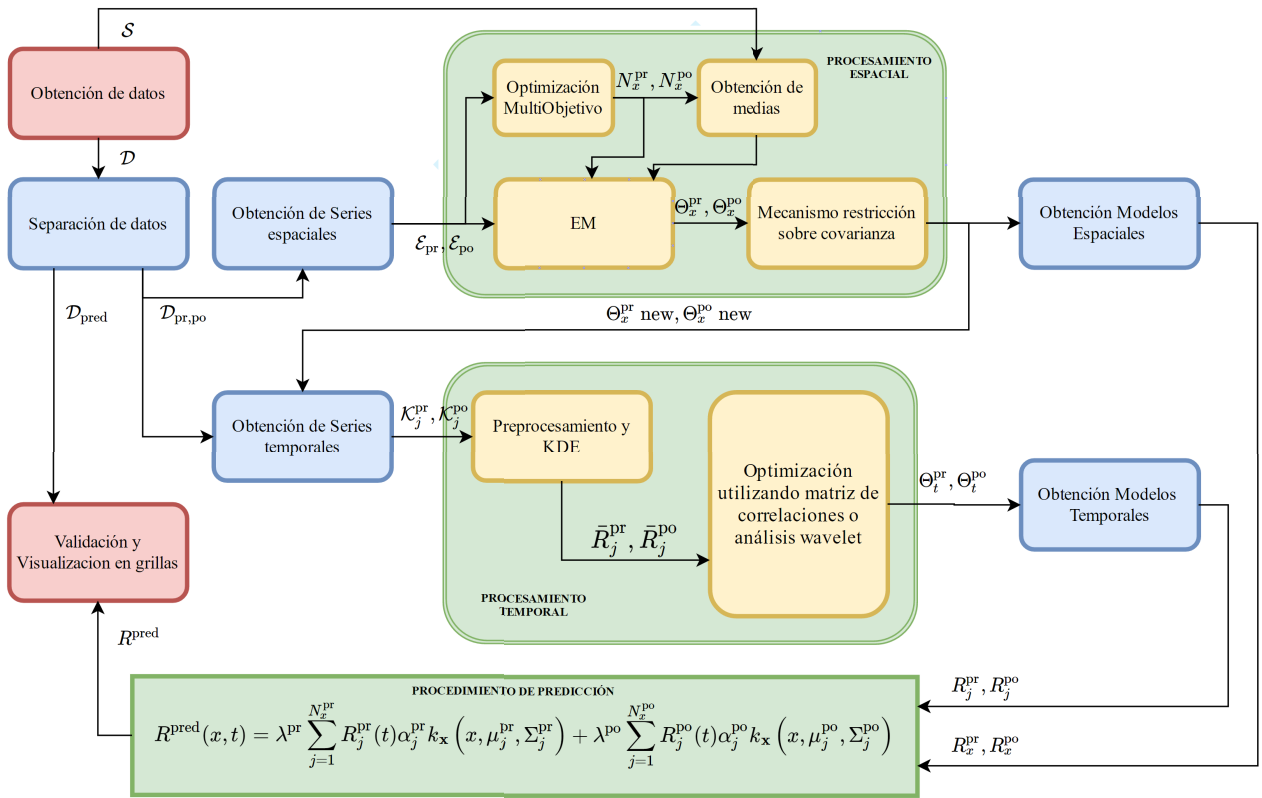


Figura 3.8: Resumen de metodología

Capítulo 4

Caso de Estudio: Modelación de Riesgo Criminal Espacio-Temporal con Datos Reales de Comunas del Gran Santiago

Este trabajo de tesis está inserto en un proyecto llevado a cabo en conjunto con CEAMOS y Carabineros de Chile, en donde el objetivo es generar un *software* que sea capaz de caracterizar el riesgo criminal para comunas de la Región Metropolitana y realizar predicciones de corto y largo plazo. La generación de predicciones y su visualización están enmarcadas en la ejecución de servidores e integración a una plataforma web llamada SAIT (Sistema de Análisis de Información Territorial). Las predicciones deben ser capaz de satisfacer las necesidades de los usuarios finales que son los comisarios y carabineros, los cuales deben trazar estrategias a nivel táctico y operativo, para ubicar recursos policiales con la debida anticipación, pero que también los retenes sepan dónde desplazarse en un determinado turno del día.

Este capítulo se centra en demostrar la efectividad del algoritmo planteado tomando como ejemplo una comuna de la Región Metropolitana. Se presentan preliminares de la metodología y luego estudios de los diferentes parámetros de la caracterización espacial considerando las diferentes alternativas de la metodología. Consecutivamente, se validan las predicciones realizadas bajo diferentes métricas y se discute sobre los diferentes problemas que conlleva medir desempeño. Además de comprobar la efectividad del algoritmo propuesto, se compara con otras metodologías que son también utilizadas en predicción de riesgo criminal. Por último se presentan resultados finales para diferentes comunas de la Región Metropolitana y se discute de por qué en algunas es más fácil y en otras más difícil predecir riesgo criminal.

El algoritmo está implementado en Python 2.7 ¹, y los resultados se visualizan ya sea mediante **matplotlib** (paquete de python) y QGIS² (herramienta de geo-visualización) y los mapas geográficos son generados en OpenStreet ³.

¹<https://www.python.org/>

²<http://www.qgis.org/es/site/>

³<http://www.openstreetmap.cl/>

4.1. Validación de Metodología Utilizando una Comuna

4.1.1. Descripción Zona de Estudio

La zona de estudio escogida corresponde a la comuna de Puente Alto, perteneciente a la Región Metropolitana y específicamente a la conurbación urbana del Gran Santiago (ver Figura 4.1), ubicada en el Sector suroriente de Santiago es la capital de la Provincia Cordillera. Posee una superficie de 88 km² siendo el 10.13 % del área del Gran Santiago. De acuerdo al Instituto Nacional de Estadísticas, en el censo de 2002, contaba con 492.915 habitantes, lo que la convertía en la comuna más poblada de Chile, inclusive el 2015 con 610.118 habitantes [18]. Al norte limita con la comuna de La Florida, al sur con la comuna de Pirque, al este con la comuna de San José de Maipo y al oeste con las comunas de La Pintana y San Bernardo.

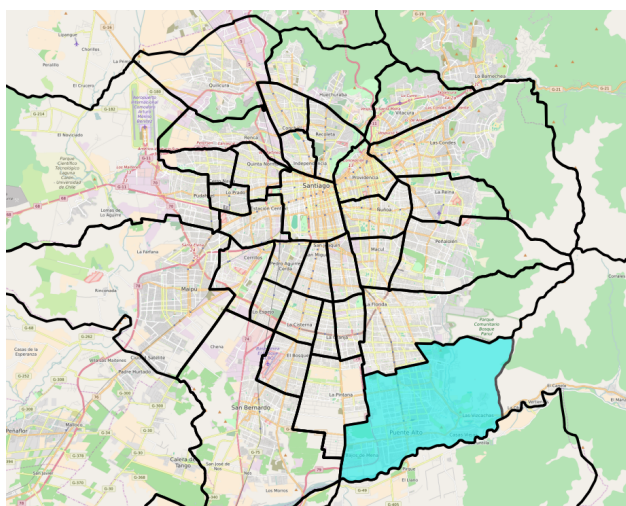


Figura 4.1: División política del Gran Santiago y comuna de Puente Alto en color celeste.

4.1.2. Base de Datos

Datos Disponibles y Datos a Utilizar

Los eventos criminales tienen una serie de atributos que los clasifican, siendo los principales el nombre del delito, código del delito, tipo de delito y su agrupado, aparte por supuesto de su localización georeferenciada, comuna de pertenencia y el registro de fecha de su ocurrencia. El código de delito es la clasificación más específica de un evento criminal, habiendo correspondencia uno a uno con el nombre del delito constituyendo aproximadamente 600 tipos diferentes. Luego, en la misma línea continúa el agrupado, que clasifica conjuntos de códigos de delitos constituyendo aproximadamente 30 tipos diferentes. Por ejemplo el agrupado **alcoholes** contiene 16 tipos de delitos siendo algunos de ellos los que se aprecian en la Tabla 4.1

AGRUPADO ALCOHOLES	CDELITO
BOTILLERIA,SUPERMERC.,MINIMARKET QUE EXP.LICOR CONSUMIDO INT.LOCAL S/P	14080
CONSUMO DE BEBIDAS ALCOHOLICAS EN LA VIA PUBLICA	14054
EBRIEDAD	14051
EXPENDIO BEBIDAS ALCOHOLICAS	14053

Tabla 4.1: Muestra de nombre de delitos y código de delitos para agrupado alcoholes.

De la misma manera están los tipos de delitos, los cuales son **Control Preventivo**, **Desordenes**, **Droga**, **DMCS** (Delitos de Mayor Connotación Social), **Infracción al tránsito**, **VIF** (Violencia Intra Familiar) y **Otros delitos**.

Para los alcances de este caso de estudio, es de interés el tipo de delito DMCS, debido a la alta victimización cercana y repetida en el tiempo de los delitos involucrados, pero además porque tienen un gran impacto en la sensación de seguridad en los habitantes. En la Tabla 5.1 de Anexos se puede apreciar los agrupados y nombre de delitos asociados al tipo DMCS.

Para efectos de este caso de estudio solo se enfocan los esfuerzos en caracterizar los agrupados correspondientes a **Robo Fuerza** y **Robo Violencia** como una sola categoría de delitos. Los demás agrupados se descartan por la poca cantidad de delitos existentes sin un patrón aparentemente inferible por medio del modelo utilizado. Para el caso de Puente Alto, se considera una base de datos histórica que parte desde 2011-01-01 hasta 2016-08-17 que contempla una distribución de datos temporales anuales hasta 2015 según la Tabla 4.2.

Año	Robo Fuerza y Robo Violencia	Robo Fuerza	Robo Violencia
2011	6795	4013	2782
2012	7272	4451	2821
2013	7004	4164	2840
2014	8333	4813	3520
2015	10271	5501	4770

Tabla 4.2: Cantidad de delitos según agrupados anual

Cabe mencionar que según las encuestas de victimización, cerca de un 50 por ciento de los delitos no son denunciados, por lo cual cualquier modelo creado a partir de esta base de datos sub-representaría el riesgo real [18]. Aún cuando se tenga este factor en contra, los patrones delictuales pueden de todas maneras caracterizarse y lograr realizar predicciones debido a su carácter cuasi-periódicos y cuasi-estacionarios. Además de los eventos criminales, se tiene conocimiento de los servicios y calles para cada comuna. En este caso particular, solo es necesaria la georeferenciación de calles y su tipo. Los servicios no son incorporados por cuanto su utilidad radica en el análisis de agrupados como hurto que por su naturaleza tienden a ocurrir en servicios. Para los agrupados de **Robo Fuerza** y **Robo Violencia**, se necesita la georeferenciación de calles para tener una idea más precisa de su ocurrencia.

División de Datos y Muestreo

Las tendencias históricas requieren de una cantidad importante de datos para dilucidar patrones. En este caso de estudio se utiliza un año de datos históricos para caracterizar el riesgo de una zona determinada. Por otro lado, las tendencias de de corto plazo debe modelar los patrones más recientes, los cuales según la literatura se pueden inferir con datos de un periodo de dos a tres meses . De esta manera, para una fecha de consulta actual f_{act} se tiene que $f_{cut} = f_{act} - 8$ semanas y $f_{ini} = f_{cut} - 1$ año. En términos de tendencia histórica puede agregarse más años de análisis, por lo cual constituye un parámetro de diseño.

Por otro lado, si bien los eventos están registrados temporalmente en formato **YYYY-MM-dd HH:mm**, estos se discretizan cada 8 horas y el índice cero se establece en f_{ini} . La razón es porque en términos operativos Carabineros de Chile dividen sus quehaceres policiales en tres turnos definidos de la siguiente manera:

- Turno 1: 07:00 - 14:59
- Turno 2: 15:00 - 22:59
- Turno 3: 23:00 - 06:59

Los turnos policiales son unidades de tiempo adecuadas para caracterizar el riesgo temporal dada por la experiencia del actuar de Carabineros de Chile, debido a que muestreos más finos incurrirían en generar más momentos con ausencia de crímenes perdiéndose capacidad para entrenar modelos, y en el caso contrario se perdería precisión por tener un muestreo grueso.

4.1.3. Resultados Preliminares

En esta sección se presentan resultados preliminares de la metodología aplicada a la comuna de estudio. Los parámetros escogidos son en base a la experiencia y heurísticas desarrolladas en la implementación del modelo, pero sin embargo son de utilidad para determinar el resultado de cada etapa de la metodología. En secciones posteriores se realizan análisis bajo el enfoque *ceteris paribus* para diferentes aspectos de la caracterización espacial y con ello determinar qué parámetros y elementos son los más influyentes en el desempeño predictivo del algoritmo, recalcando que el procedimiento temporal es fijo utilizándose reconstrucción mediante Wavelet y proyección de señales pretéritas (Sección 3.4.1). Los siguientes resultados se obtienen considerando eventos de la unión de los agrupados **Robo Fuerza** y **Robo Violencia**, considerando también las fechas $f_{act} = 2016-08-01$, $f_{cut} = 2016-06-01$ y $f_{ini} = 2015-06-01$.

Caracterización Espacial

Los parámetros para generar una GMM son ejecutados mediante el paquete de Python **sklearn.mixture.GMM**, para el caso *a priori* y *a posteriori* se utilizan los parámetros fijados de manera heurística según la Tabla 4.3.

	<i>a priori</i>	<i>a posteriori</i>
<i>n_components</i>	179	69
<i>covariance_type</i>	“full”	“full”
<i>min_covar</i>	1e-8	1e-8
<i>params</i>	wmc	wmc
<i>n_iter</i>	1000	1000

Tabla 4.3: Parámetros para caracterización *a priori* y *a posteriori* típicas, donde *n_components* es la cantidad de componentes de la GMM, *covariance_type* indica el tipo de covarianza siendo en este caso con vectores propios arbitrarios, *min_covar* es el mínimo valor de la covarianza diagonal, *params* corresponde a los parámetros a entrenar siendo *weights*, *means* y *covariances* y *n_iter* es la cantidad de iteraciones del algoritmo EM.

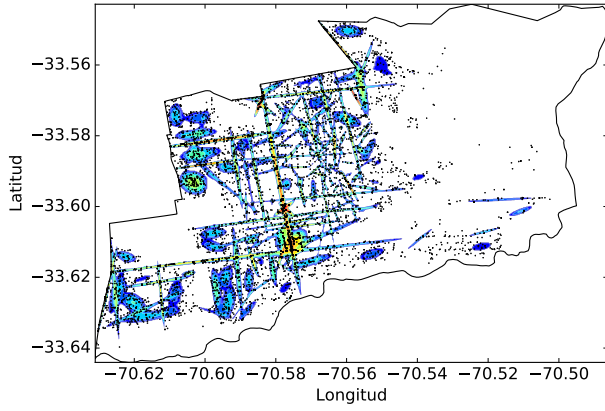
De esta manera se obtienen las caracterizaciones espaciales *a priori* (ver Figura 4.2a) y *a posteriori* (ver Figura 4.2c) donde se computa el logaritmo de la densidad de riesgo (tonos rojos, verdes y azules) y los puntos negros corresponden a los eventos criminales. Además, se aprecia el resultado de las caracterizaciones de manera tridimensional (ver Figura 4.2b y 4.2d) donde el eje Z es la densidad de riesgo.

Caracterización Temporal

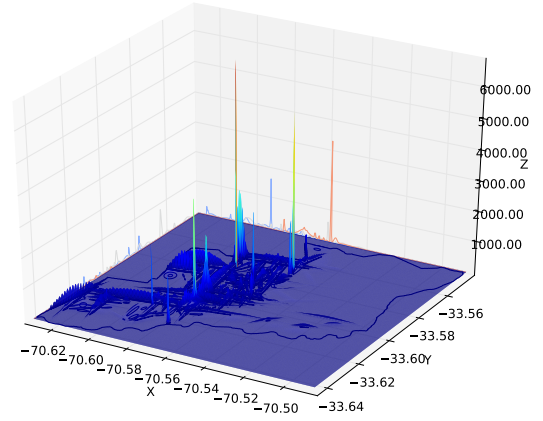
En este punto se construyen los conjuntos $\mathcal{K}_j^{\text{pr}} \forall j = 1, \dots, 179$ y $\mathcal{K}_k^{\text{po}} \forall k = 1, \dots, 69$ (ver (2.21)) que corresponden a los eventos asignados a cada componente espacial *a priori* y *a posteriori* respectivamente. A modo de ejemplo, en la caracterización *a priori*, se tienen las asignaciones de los eventos criminales con sus respectivos centros rojos como se aprecia en la Figura 4.3a (líneas que unen cada evento con su respectivo componente), asimismo se obtiene el histograma de la cantidad de eventos por componente en la Figura 4.3b, indicando que en promedio se tiene 50 delitos por kernel. En secciones posteriores se analiza en detalle cuál debiese ser el número óptimo de componentes según los criterios de *Silhouette*+Aciertos y *Silhouette*+Desviaciones contrastados mediante el desempeño predictivo del algoritmo. Se constata además, la existencia de componentes que correlacionan eventos criminales muy lejanos entre sí, lo que en términos fenomenológicos no tiene sentido, por lo que en secciones posteriores se analiza en detalle el procedimiento resumido de la Figura 3.3.

Algunas veces ciertos kernels poseen una baja cantidad de eventos criminales asignados (menos de 10 por kernel⁴), por lo cual no es posible realizar un entrenamiento temporal adecuado (generar una serie de tiempo a partir de los eventos asignados). En esos casos, se fija la ponderación temporal en $R_j(t) = 1$ considerándose solo el efecto espacial (ver (3.37)). El argumento de realizar el procedimiento anterior estriba en que el peso de un kernel con baja cantidad de delitos se puede caracterizar sin problemas espacialmente, por lo cual no es necesario agregar un factor dinámico. En caso de agregarlo se puede estar distorsionando el riesgo asociado al kernel estudiado debido a que no hay una correcta estimación de los componentes de frecuencia de la serie de tiempo asociada. Además en el caso que existan kernels sin delitos asignados, se descartan del procedimiento temporal y del de predicción.

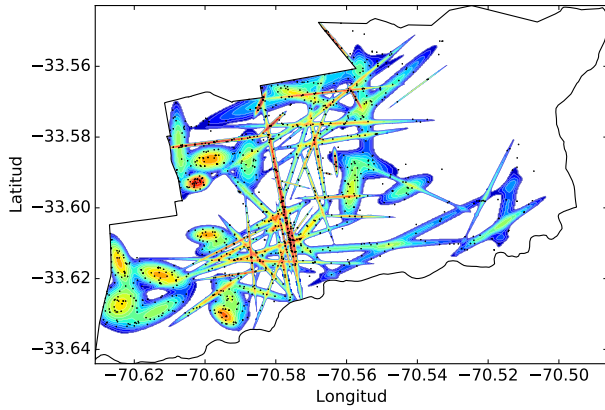
⁴el umbral igual a 10 se fija por experiencia



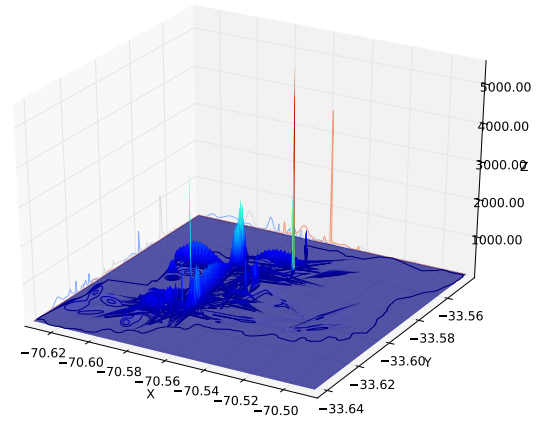
(a) *a priori* con 20% de cobertura



(b) *a priori* 3D

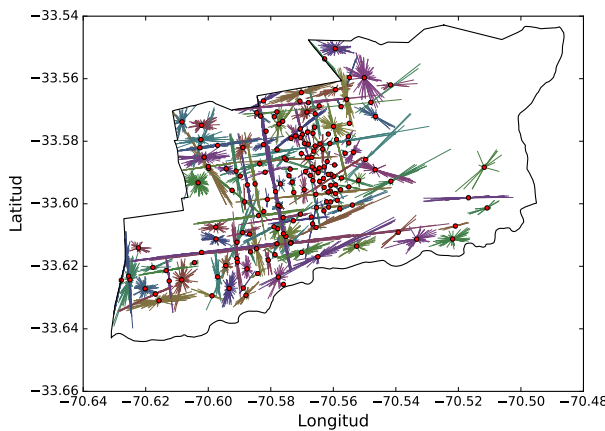


(c) *a posteriori* con 40% de cobertura

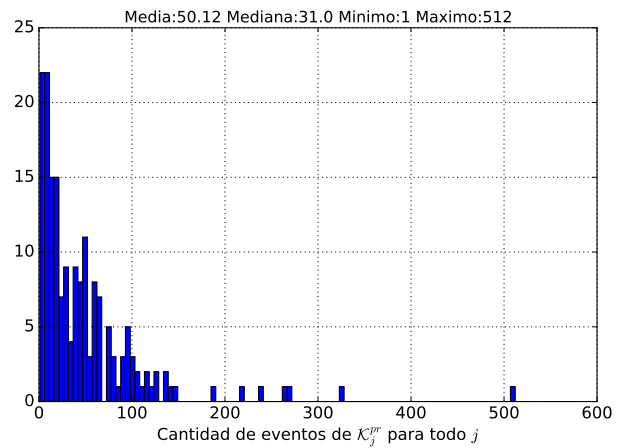


(d) *a posteriori* 3D

Figura 4.2: Caracterización espacial de modelos.



(a) Pertenencia eventos



(b) Cantidad de eventos por kernels *a priori*

Figura 4.3: Pertenencia de eventos \mathcal{E}_{pr} a $\mathcal{K}_j^{\text{pr}} \forall j = 1, \dots, 179$ (considerando centroides en color rojo) y construcción de histograma de cantidad de eventos de cada $\mathcal{K}_j^{\text{pr}}$.

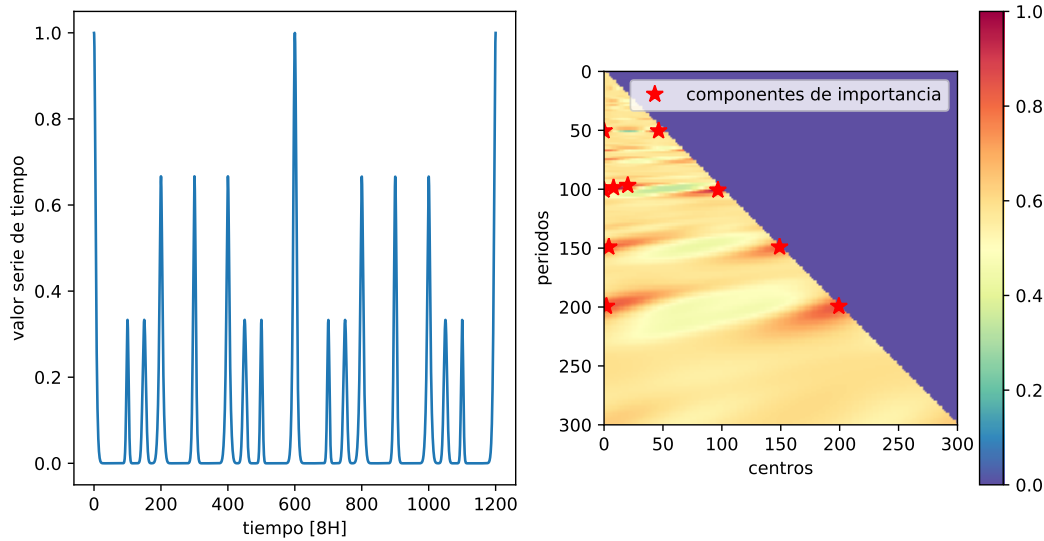


Figura 4.4: Serie de tiempo de ejemplo suavizada con KDE con su respectiva matriz de correlaciones destacando los primeros 10 pares de centros y periodos (en unidades de 8 horas)

En cuanto a la reconstrucción de series de tiempo, si se realiza por medio del entrenamiento de un modelo por mezcla de kernels periódicos, en primer lugar se debe construir la matriz de correlaciones como se aprecia en la Figura 4.4. Con esta matriz de correlaciones se seleccionan los 10 centros y periodos más importantes (que tienen una mayor correlación), de esta manera si el espacio paramétrico para cada serie temporal está definido por:

$$\Theta_R = \{(\delta_i, t_i, l_i, p_i)\}_{i=1}^{N_t} \quad (4.1)$$

solo basta entrenar los ponderadores lineales δ_i y los ancho de banda l_i . Si se desea simplificar más aún la optimización el ancho de banda se deja fijo en aproximadamente 4 horas, por lo que solo se deben entrenar los ponderadores lineales de cada kernel periódico.

Por otro lado, la reconstrucción de las series de tiempo también se puede realizar por medio de la Transformada Continua Wavelet como se aprecia en la Figura 4.5, donde se visualiza otro ejemplo de serie de tiempo suavizada mediante el método KDE y su reconstrucción tomando las 10 escalas más importantes del Espectro de Energía Wavelet (siguiendo el procedimiento de la Sección 3.3.3). Para la reconstrucción se ocupa el paquete de python **mlpy.wavelet** donde se utiliza la función base “DOG”, la fracción de resolución ocupada es $8/24.0$ (considerando que cada unidad de tiempo son 8 horas) y la resolución de las escalas es de 0.025 . Además se itera diferentes valores de m de la función base siendo $[2,6,10]$ y el ancho de las bandas del histograma de energía (que se aprecia en la Figura 3.7) se varía entre $[1,2,3,5]$. El valor óptimo de m y ancho de las bandas del histograma de energía se calcula buscando la reconstrucción de la serie de tiempo mediante la Antitransformada Wavelet minimizando el error cuadrático medio (considerando las 10 escalas de mayor energía).

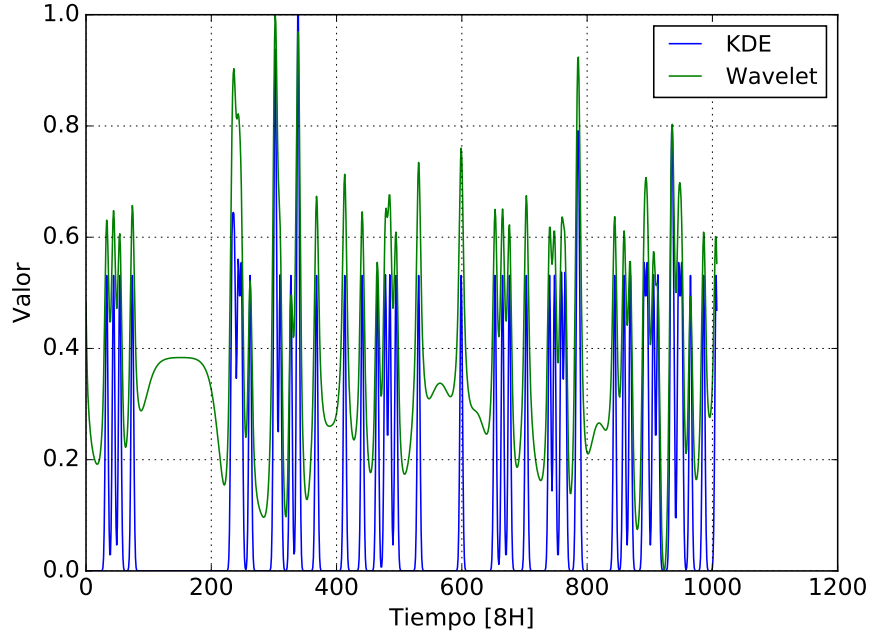


Figura 4.5: Serie de tiempo suavizada con KDE de un kernel de $\mathcal{K}_j^{\text{Pr}}$ y su versión reconstruida mediante Transformada Wavelet (ver Sección 3.3.3).

Predicción

Considerando las señales temporales reconstruidas, aquellas que pertenecen a la clasificación *a priori* su predicción se realiza considerando $R_j^{\text{Pr}}(t_{\text{pred}}) = R_j^{\text{Pr}}(t_{\text{pred}} - t(1 \text{ año}))$ y aquellas que se clasifican como *a posteriori* su predicción se realiza mediante el procedimiento de Proyección de Señales Pretéritas de la Sección 3.4.1 considerando una periodicidad $p_2 = 1$ semana y factor de olvido $\gamma^{\text{Po}} = 0.9$. La elección de parámetros anteriores permite una predicción no más allá de una semana, por lo cual con una resolución de 8 horas, solo puede realizarse una predicción de a lo más 21 pasos.

Además se escoge para simplificar el proceso de predicción se escoge el $\lambda^{\text{Pr}} = \lambda^{\text{Po}} = 0.5$ (ver Ec (3.37)) y se generan 4 pasos de predicción como se aprecia en la Figura 4.6.

Como se puede observar, las predicciones son casi 100 % certeras considerando una cobertura del 30 % del área total de la comuna, sin embargo puede resultar ser mucha área para la cantidad de recursos policiales disponibles. Es por ello que la métrica AUC en conjunto de las curvas HRP y TIP analizan el desempeño predictivo para diferentes niveles de cobertura como se presenta a continuación.

Métricas de Predicción

En este caso se calcula la métrica AUC y las curvas HRP y TIP condicionando las predicciones a la grilla de estudio aportada por Carabineros de Chile que se aprecia en la Figura 4.7

En cada grilla se integra la superficie de riesgo obtenida y se calculan las métricas ya

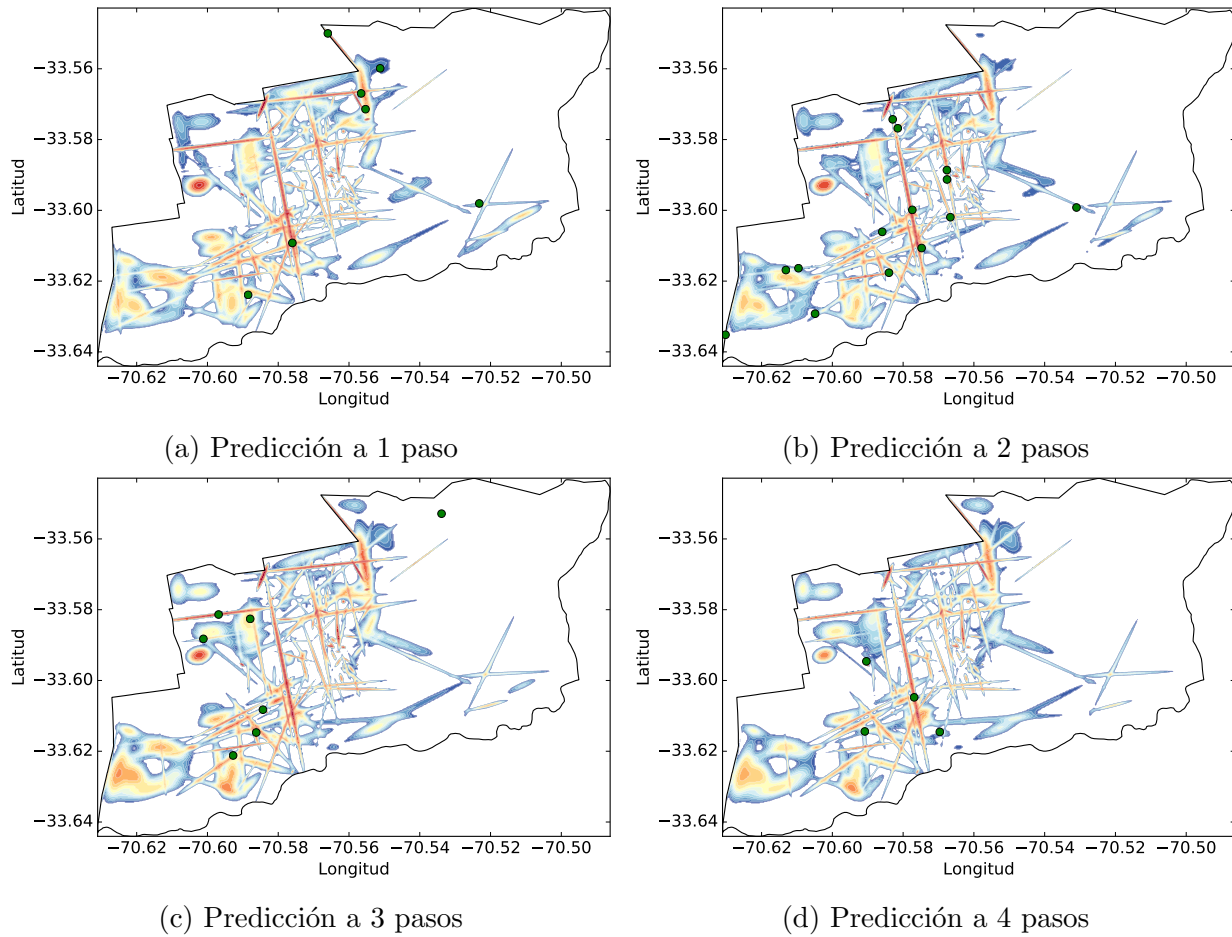


Figura 4.6: Predicciones de 1 a 4 pasos considerando un 30% del área de cobertura total de la comuna.

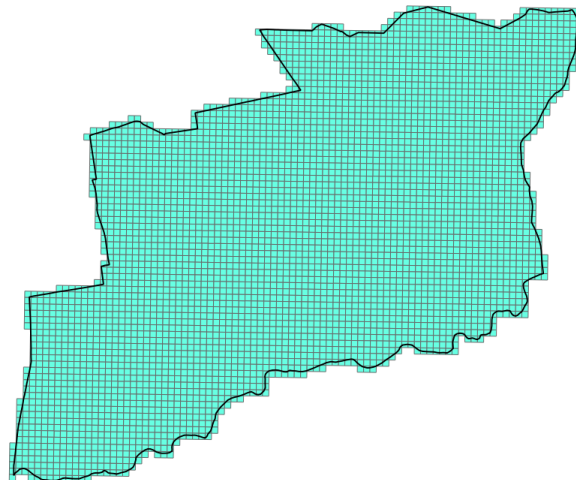
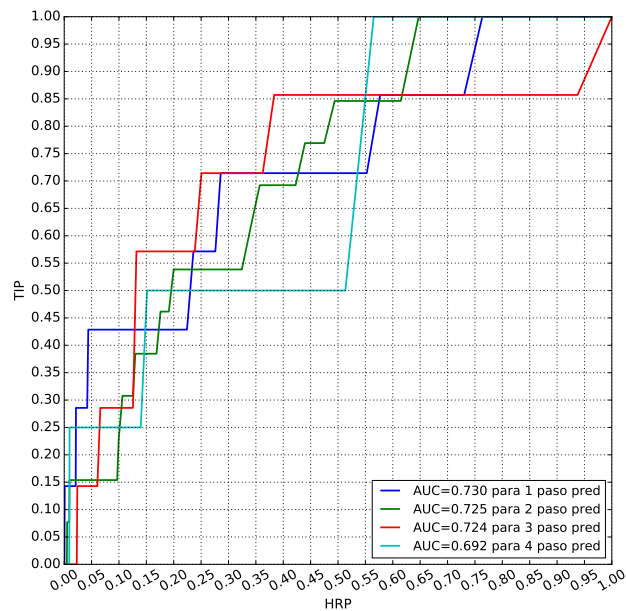


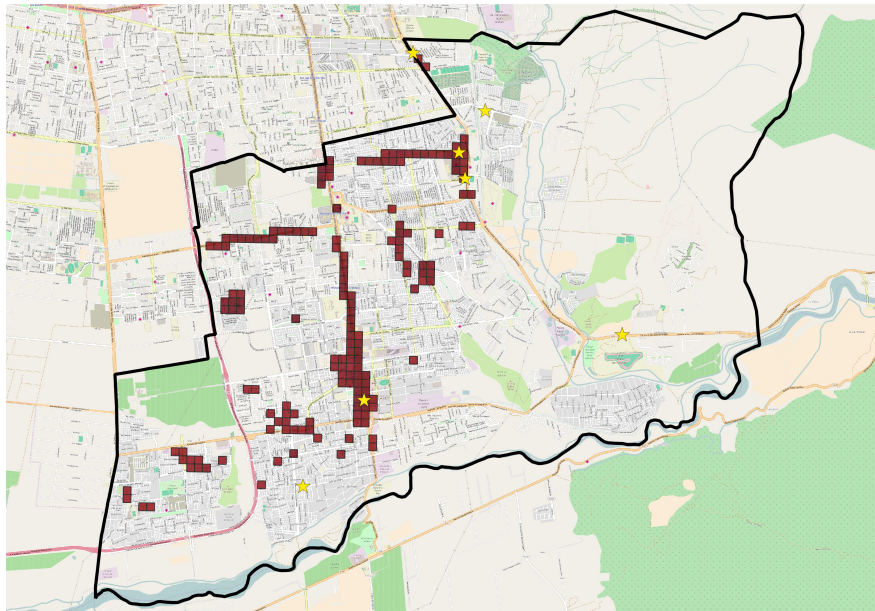
Figura 4.7: Grillas de estudio para comuna de Puente Alto.

descritas como se aprecia en la Figura 4.8a, destacándose que para un paso de predicción pintando aproximadamente el 5% se tiene un 42.8% y que además el AUC es mayor, degradándose en cada paso de predicción. Un ejemplo es llevado a cabo con las grillas de estudio

como se aprecia en la Figura 4.8b.



(a) Curvas HRP y TIP



(b) Mapa con grillas

Figura 4.8: Curva HRP v/s TIP para 4 pasos de predicción con el respectivo valor AUC y un ejemplo para 1 paso de predicción mostrando 164 grillas de mayor riesgo (3.9%) con acierto de 3 de 7 delitos (42.8%).

Cabe mencionar que para validar la metodología estadísticamente se debe hacer un análisis exhaustivo de casos para diferentes fechas, pasos de predicción y comunas, puesto que para una resolución de 8 horas, los delitos de Robo con Fuerza y Violencia no son numerosos para cada paso de predicción.

En las siguientes secciones se realiza el análisis *ceteris paribus* para diferentes aspectos del

algoritmo siendo la caracterización espacial del método el foco principal.

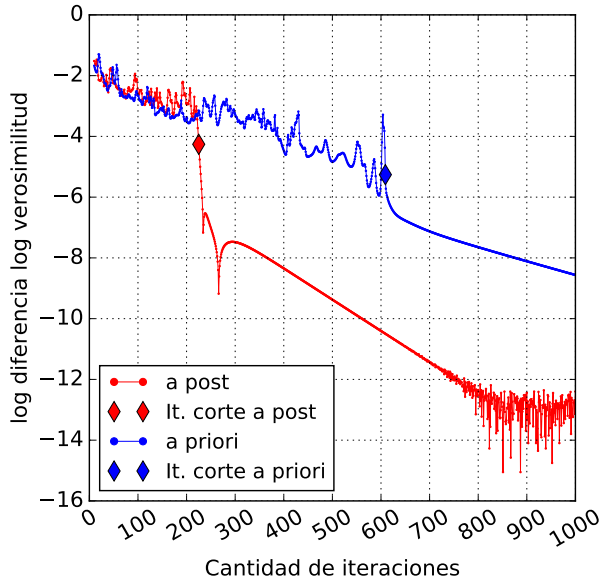
4.1.4. Estudio de Criterios de Término de Iteraciones en Modelos Espaciales

En el **estudio de cantidad de componentes iniciales en modelos espaciales** se asume que la cantidad de iteraciones realizadas mediante EM son las adecuadas, vale decir que los parámetros calculados no ofrecen cambios sustanciales entre iteraciones. Un criterio usual es la **tolerancia**, que se calcula como la diferencia del logaritmo de la verosimilitud entre iteraciones consecutivas. Lamentablemente con el criterio de tolerancia no es fácil establecer un valor fijo, puesto que no es intuitivo diferenciar por ejemplo una tolerancia de $1e-5$ y $1e-7$. Es por ello que este estudio se centra en analizar para una cantidad fija de componentes, cuáles deben ser las iteraciones adecuadas en modelos *a priori* como *a posteriori*. Cabe mencionar que no realizar una cantidad de iteraciones adecuadas puede perturbar la elección óptima de componentes espaciales debido a que las funciones de mérito se distorsionarían. Un claro ejemplo es considerar una cantidad insuficiente de iteraciones obteniendo un modelo con componentes muy aglutinados, por lo cual la métrica *Silhouette* no arrojaría resultados adecuados. Para tener indicios sobre la convergencia de medias, covarianzas y pesos, que son los parámetros del entrenamiento de una GMM se introduce el criterio de **distancias**:

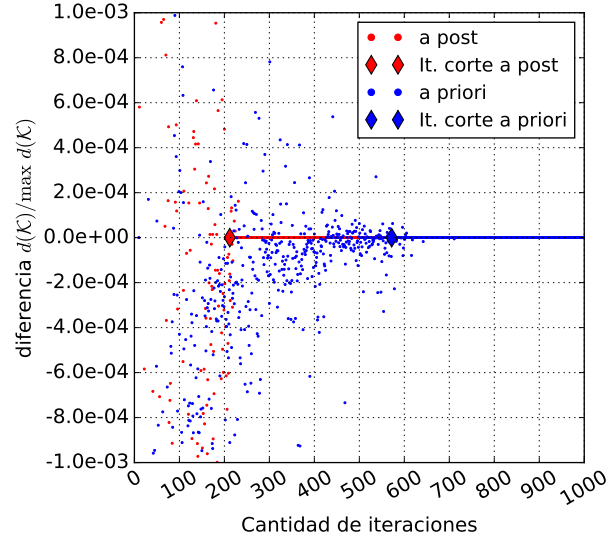
$$d(\mathcal{K}) = \sqrt{\sum_{k \in \mathcal{K}} |k|^2} \quad (4.2)$$

donde \mathcal{K} es la colección de conjuntos de datos asignados a cada kernel espacial para un modelo GMM. Mediante $d(\mathcal{K})$ se identifica el momento en que las clasificaciones de datos con sus respectivos kernels se estabilizan. En la Figura 4.9 se computa para los modelos *a priori* y *a posteriori* de la sección anterior y basado en el criterio de **tolerancia**, el criterio de **diferencia de tolerancias**, el cual se visualiza en escala logarítmica para mayor claridad.

Mediante este criterio, se selecciona la iteración de corte como aquella en que la tolerancia haya experimentado una variación de más de **dos órdenes de magnitud en descenso** (dos unidades en escala logarítmica) en **al menos 10 iteraciones consecutivas** del algoritmo EM. Mediante esta heurística la selección de la iteración de corte para el modelo *a priori* resulta ser adecuada porque a partir de la iteración 600 la ganancia en verosimilitud es marginal. Para el caso *a posteriori*, el algoritmo se detiene en las 240 iteraciones aproximadamente, sin embargo a juzgar por las iteraciones siguientes, la ganancia tiene un descenso brusco con dos mínimos locales, por lo cual, pese al criterio de detención, no se tiene seguridad de que los parámetros se hayan estabilizado. Por otra parte, se analiza el criterio de distancia según (4.2) como se aprecia en la Figura 4.9b. Los diamantes indican las iteraciones de corte, que para el caso *a priori* es en las 572 iteraciones y en el caso *a posteriori* es en las 212. Las iteraciones de corte se calculan cuando la **cantidad de datos en cada kernel (clasificaciones) no varía en a lo más 10 iteraciones**.



(a) criterio tolerancia



(b) criterio distancia

Figura 4.9: Iteraciones de modelos espaciales *a priori* y *a posteriori* con 179 y 69 componentes respectivamente computando criterio de tolerancia y distancia para diferente cantidad de iteraciones.

Discusión

El criterio de distancias no es costoso de computar en cada iteración, por lo cual es factible considerarlo como criterio *online* de detención. Sin embargo, de todas maneras se deben realizar cálculos exhaustivos considerando zonas de estudios donde la cantidad de kernels para las GMM's sea mucho mayor, esto para comprobar que su costo computacional no es tan elevado como para volver a criterios usuales tales como **tolerancia** o simplemente fijar una cantidad de iteraciones máxima.

4.1.5. Estudio Cantidad de Componentes Iniciales en Modelos Espaciales

En este análisis se estudia la generación de modelos espaciales con diferente cantidad de componentes considerando que la condición de término se ciñe por el criterio de convergencia de distancias de la sección anterior. Para este caso se trabaja con dos optimizaciones multi-objetivo siendo *Silhouette* con *Aciertos* y *Silhouette* con *Desviaciones* las cuales ilustran desde un punto de vista espacial la calidad de la caracterización. Estas métricas son costosas de calcular por lo que no se ocupan para tomar decisiones instantáneas en una determinada caracterización, pero sí, para establecer criterios para tener una idea más clara de cuántos componentes espaciales son razonables a utilizar sin realizar un estudio exhaustivo. Siguiendo con la configuración del estudio anterior y la sección de resultados previos, se estudia la caracterización *a priori* y *a posteriori* considerando diferente cantidad de componentes espaciales para los criterios expuestos como se aprecia en la Figura 4.10. Debido a que no

hay claridad de qué ponderación sobre los criterios es más adecuada, se grafican varias ponderaciones donde el color rojo indica 100% *Silhouette* y el azul indica 100% ya sea Aciertos o Desviaciones según corresponda.

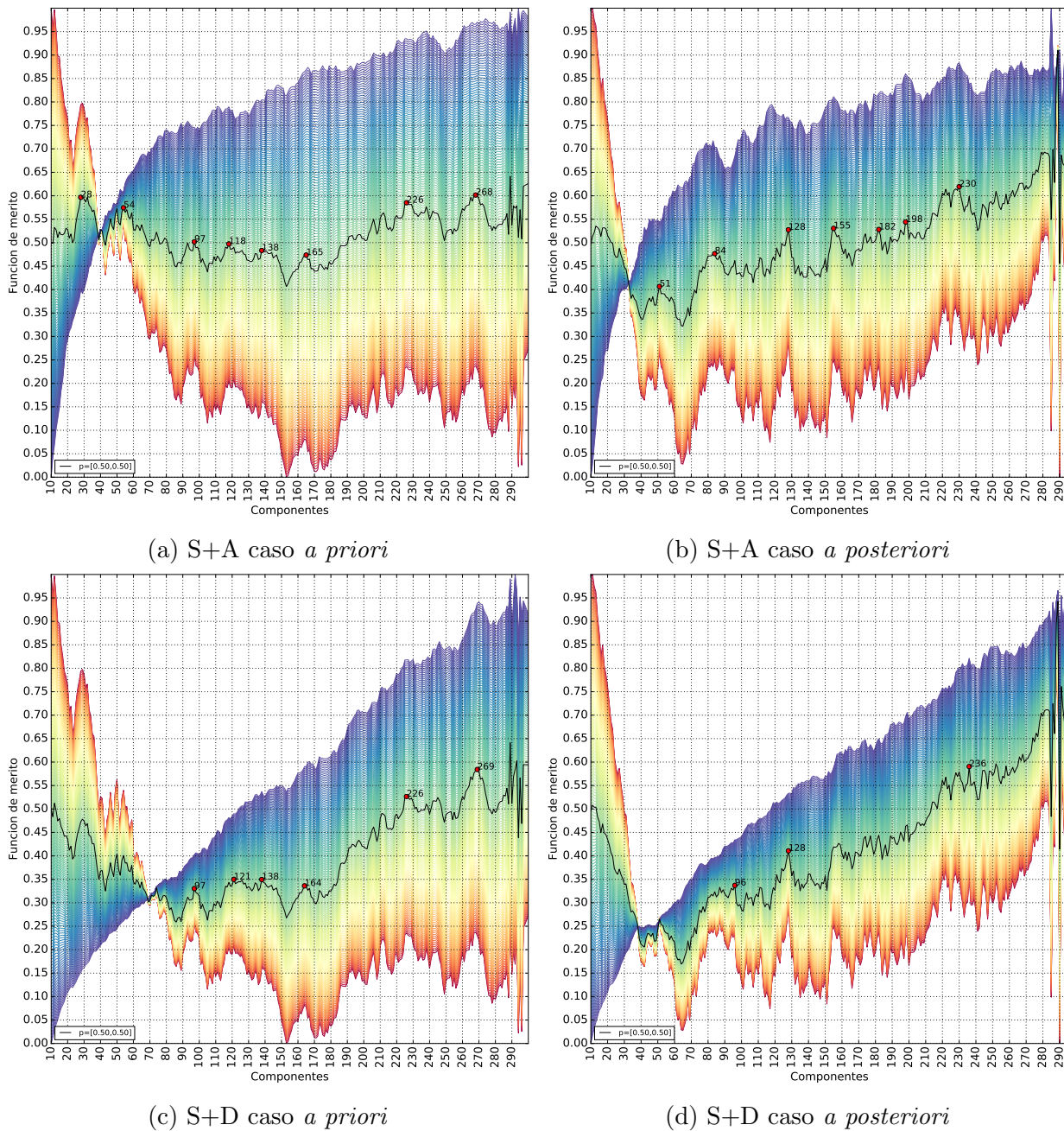


Figura 4.10: Criterios multi-objetivo (*Silhouette* con Aciertos y *Silhouette* con Desviaciones) en función de la de cantidad de componentes espaciales para modelo *a priori* y *a posteriori*, utilizando múltiples ponderaciones donde [1.0,0.0] indica 100% *Silhouette* (tono rojo) y [0.0,1.0] indica 100% Aciertos o Desviaciones según corresponda (tono azul). Se selecciona máximos locales con una ventana de tamaño 11 para una ponderación de [0.5,0.5] (línea negra).

Como se aprecia en la Figura 4.10a para el caso *a priori* *Silhouette* con Aciertos tiene un nodo en 40 componentes indicando un punto de equilibrio independiente de las ponderaciones.

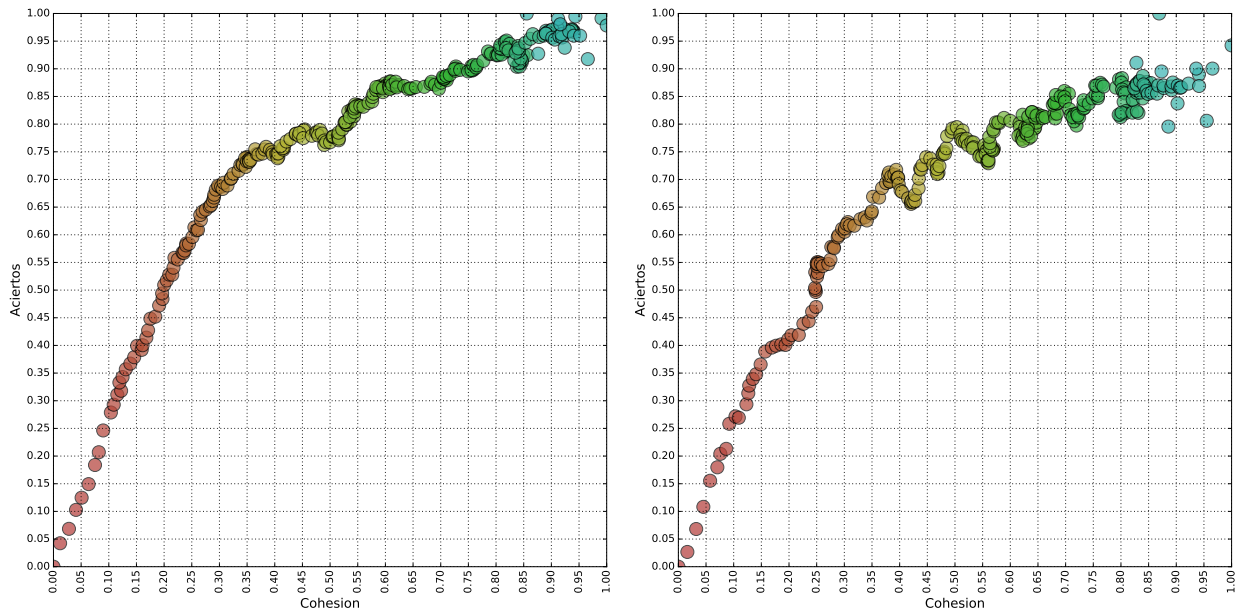
Como se debe maximizar la función de mérito para una determinada ponderación, si se considera $p = [1.0, 0.0]$ se privilegiaría *Silhouette* ante Aciertos tomando solo 10 componentes, mientras que para el caso contrario se estarían escogiendo 300 componentes. Es por ello que se toma la ponderación $p = [0.5, 0.5]$ (con previa suavización y normalización de las curvas) y se seleccionan máximos locales (dentro de una ventana de tamaño 11), esto para no perjudicar algún posible candidato a óptimo que pudo haber salido en otra ponderación. El mismo procedimiento se realiza para las demás Figuras: donde para el caso *a posteriori* de la Figura 4.10b el nodo se encuentra en el componente 33 que es menor que el del caso *a priori* anterior (habiendo concordancia con la cantidad de datos en la caracterización utilizada en los modelos). Para este caso los máximos locales encontrados son 51,84,128,155,182,198 y 230, recalcando que a diferencia del caso anterior *Silhouette* tiende a aumentar a partir de las 100 componentes. Por otra parte, para los criterios *Silhouette* con Desviaciones⁵ del modelo *a priori* como se aprecia en la Figura 4.10c el nodo se desplaza a 70 componentes mientras que para el caso *a posteriori* de la Figura 4.10d se desplaza de 33 a 37. Cabe mencionar que en este último caso tanto *Silhouette* y Desviaciones muestran una tendencia ascendente a partir de las 170 componentes.

Comparando S+A y S+D, se puede notar la semejanza de los máximos locales escogidos. Este fenómeno no es extraño, puesto que tanto el criterio de Desviaciones como Aciertos son maximizados en la medida que aumenta la cantidad de componentes, exhibiendo una tasa de crecimiento diferente como se evidencia en la Figura 4.11, en donde la cantidad de componentes aumenta de izquierda a derecha y de abajo hacia arriba (de 10 a 300 componentes). Si ambos criterios tuviesen la misma tasa de crecimiento entonces constituirían una línea recta que parte del origen y pendiente uno, pero como se aprecia en el caso *a priori* (Figura 4.11a) la curva tiene un punto de inflexión cuando el criterio de Aciertos está en 0.75, creciendo con una pendiente más pequeña. Ahora considerando el caso *a posteriori* (Figura 4.11b), se tiene un comportamiento similar, sin embargo el último tramo a partir de 0.4 del criterio de Cohesión existen perturbaciones.

Si bien Desviaciones y Aciertos apuntan en una misma dirección al maximizar, se enfocan en aspectos diferentes en la caracterización. En caso de privilegiarse la homogeneidad de la cantidad de eventos criminales de los conjuntos $\mathcal{K}_j^{\text{pr}}$ y $\mathcal{K}_j^{\text{po}}$ entonces la función de mérito *Silhouette* + Desviaciones debe primar, en cambio, si se privilegia la cantidad de eventos criminales acertados en el entrenamiento (dada una cantidad fija de área a pintar que en este caso es del 10%), entonces la función de mérito *Silhouette* + Aciertos debe ser la adecuada.

La optimización multi-objetivo tiene como fin entregar cuáles son componentes espaciales óptimos N_x^{pr} y N_x^{po} de acuerdo a criterios meramente espaciales, sin embargo, el objetivo final es establecer los componentes óptimos de acuerdo al desempeño predictivo de la metodología. Es por ello, que para validar y discernir qué máximos locales resultan ser más adecuados a partir de la optimización espacial, se utiliza la métrica predictiva AUC. Debido a que no se tiene claridad de la combinación ideal de máximos locales *a priori* ni *a posteriori*, se comprueba la capacidad predictiva para todas las combinaciones considerando lo que dice el criterio S+A y S+D juntos. Para este análisis se calcula el valor del AUC promedio considerando 3 pasos de predicción y también se visualiza el resultado de los tres (ver Figura 4.12) con el objetivo de verificar el impacto predictivo de una elección de componentes con respecto a

⁵En realidad es el criterio de Cohesión explicado en la Sección 2.5.4



(a) Cohesión versus Aciertos caso *a priori* (b) Cohesión versus Aciertos caso *a posteriori*

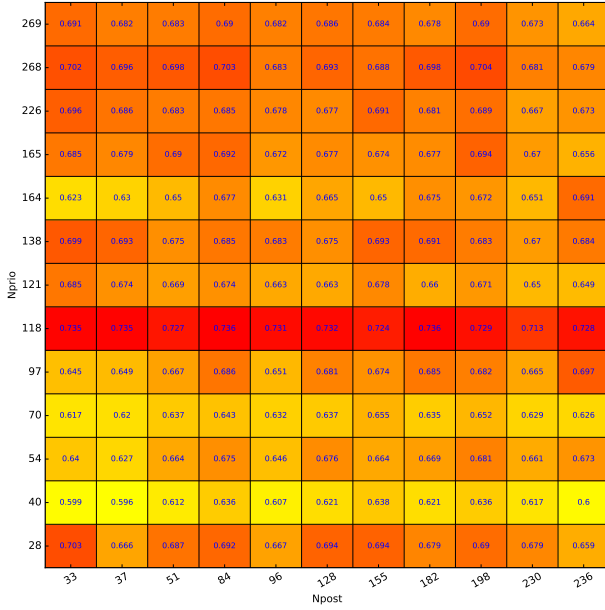
Figura 4.11: Relación entre criterio de Desviaciones (Cohesión) y Aciertos para modelo *a priori* y *a posteriori* considerando 10 a 300 componentes espaciales correspondiente a los puntos que parten desde el borde inferior izquierdo y terminan en el borde superior derecho.

otra.

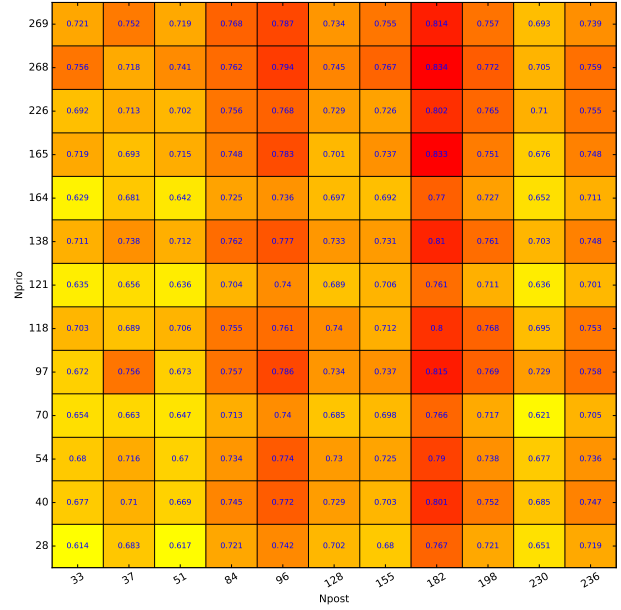
Como se aprecia en la Figura 4.12a para la predicción a un paso, utilizando 118 componentes *a priori* se obtienen los valores de AUC más altos casi independiente de la cantidad de componentes *a posteriori*, sin embargo, se tiende a privilegiar el componente 84 seguido del 37 y 33 que corresponden a nodos. Ahora bien, en la Figura 4.12d para una predicción promedio de tres pasos las predicciones son mejores cuando se escoge una cantidad de componentes *a priori* más altas, por ejemplo 269 y para el caso *a posteriori* siguiendo la misma tendencia se escoge por ejemplo 182.

Discusión

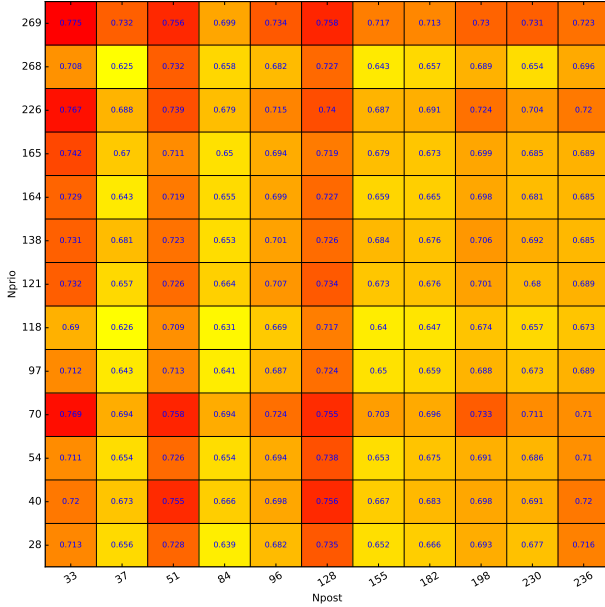
Las optimizaciones multi-objetivo presentadas, dan luces de cuáles son las cantidades óptimas de componentes a utilizar en los modelos *a priori* y *a posteriori* estudiados. Sin embargo, no son valores que se puedan determinar con certeza puesto que para diferentes pasos de predicción se obtienen diferente cantidad de óptimos. En términos de costo computacional no es conveniente calcular un modelo por cada paso de predicción, es por ello que es mejor trabajar con el promedio de los tres pasos de predicción para calcular un determinado óptimo. La pregunta que debe responderse es si mediante estas optimizaciones multi-objetivo (que son aplicadas solo a nivel espacial) existe alguna correlación con los resultados predictivos presentados. Como se evidencia en la Figura 4.12d, la concentración de mayores AUC's se encuentran a la derecha de los nodos de la Figura 4.10, es decir, para el caso *a priori* deben haber al menos 70 componentes, y en el caso *a posteriori* deben existir al menos 51. Asimismo,



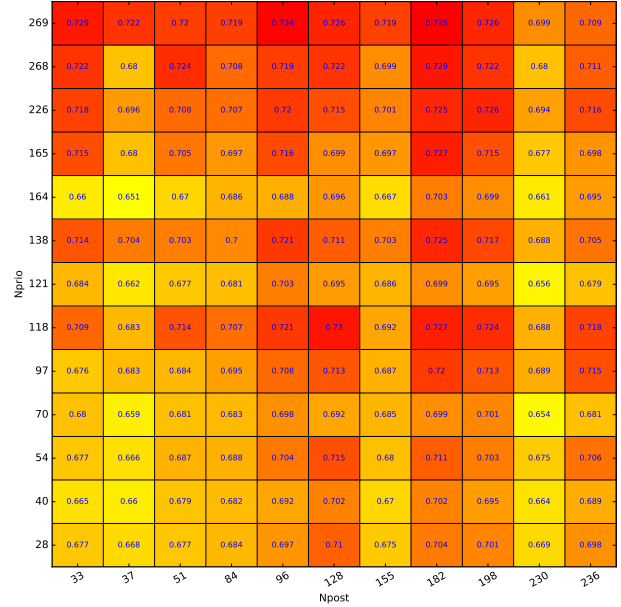
(a) AUC predicción a un paso.



(b) AUC predicción a dos pasos.



(c) AUC predicción a tres pasos.



(d) AUC promedio de los tres pasos.

Figura 4.12: Valores de AUC a un paso y promedio de tres pasos para diferentes componentes espaciales de modelo *a priori* y *a posteriori*.

mo, los óptimos no se concentran utilizando la máxima cantidad de componentes en ambos modelos ($N_{\text{prio}} = 269$, $N_{\text{post}} = 236$), pero sí, se destaca que al menos se deben ocupar más componentes *a priori* que *a posteriori* ($N_{\text{prio}} = 269$, $N_{\text{post}} = 96$ que resultan ser el último máximo local escogido por S+D en caso *a priori* y el primer máximo local a la derecha del nodo para S+D del caso *a posteriori*). Por último, para poder determinar con mayor certeza un criterio adecuado espacial mediante estas optimizaciones multi-objetivo se deben computar exhaustivamente pruebas para variadas ventanas de datos, con fechas y zonas de estudio diferentes.

4.1.6. Estudio de Mecanismo sobre Valores Propios de Matrices de Covarianza

Este estudio se enfoca en analizar cómo varían las caracterizaciones espaciales si dada una cantidad de componentes fijas espaciales se aplica el **mecanismo para cumplir con restricción sobre valores propios** de la Sección 3.2.2. Para activar el mecanismo sobre una componente espacial se debe cumplir que la razón entre el mayor valor propio con respecto al menor supere el umbral r^2 . De la aplicación del mecanismo se desprenden dos parámetros de diseño que corresponden a:

- Distancia de los nuevos componentes con respecto al centro: γ · eje mayor
- Largo del eje mayor de las nuevas componentes: β · eje menor

El objetivo de este procedimiento es eliminar las correlaciones espaciales de eventos criminales que están muy lejanos geográficamente. En ese sentido la apuesta es mejorar el desempeño predictivo para que la caracterización tanto espacial como temporal representen el comportamiento local de una zona y que no se vea distorsionadas por otras. Para efecto de análisis, el mecanismo se ejecuta solo una vez a las componentes que no cumplan con la restricción (ya que podría hacerse una iteración recursiva sobre las nuevas componentes). Sin pérdida de generalidad los parámetros se fijan de tal manera que los ejes mayores de las nuevas elipses sean un tercio del tamaño de la original, mientras que sus ejes menores sean igual a la original. De esta manera solo se varía r^2 y γ . A modo de ejemplo, en la Figura 4.13 se presenta la ejecución del mecanismo sobre el modelo *a posteriori*, el cual en primera instancia tiene 40 componentes (Figura 4.13a) y utilizando $r^2 = 6$ y $\gamma = 0.5$, 21 componentes de las 40 no cumplen con la restricción y por tanto son modificadas como se ha explicado (ver Figura 4.13b) teniendo ahora en total 82 componentes.

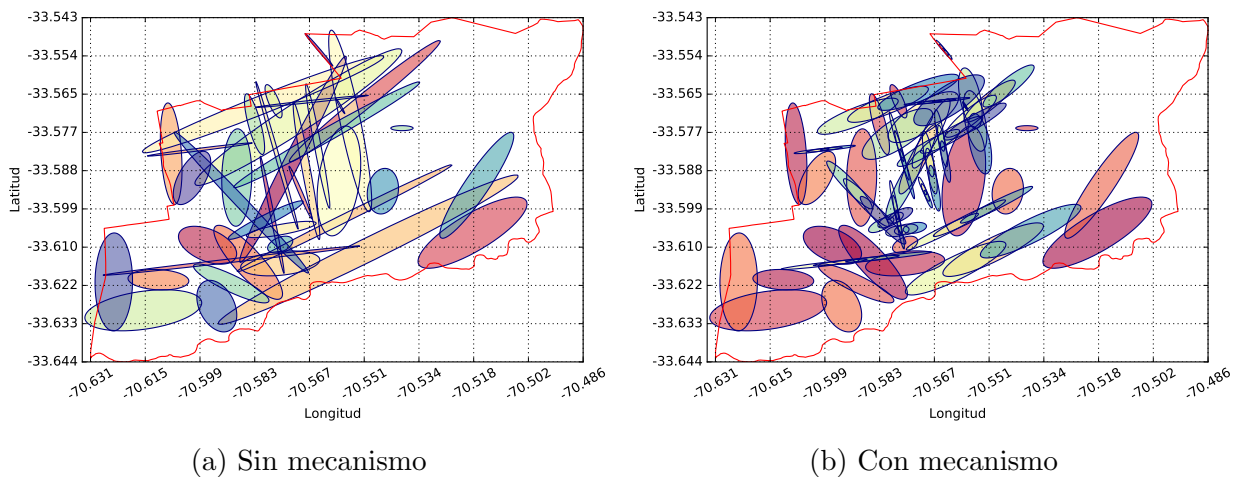
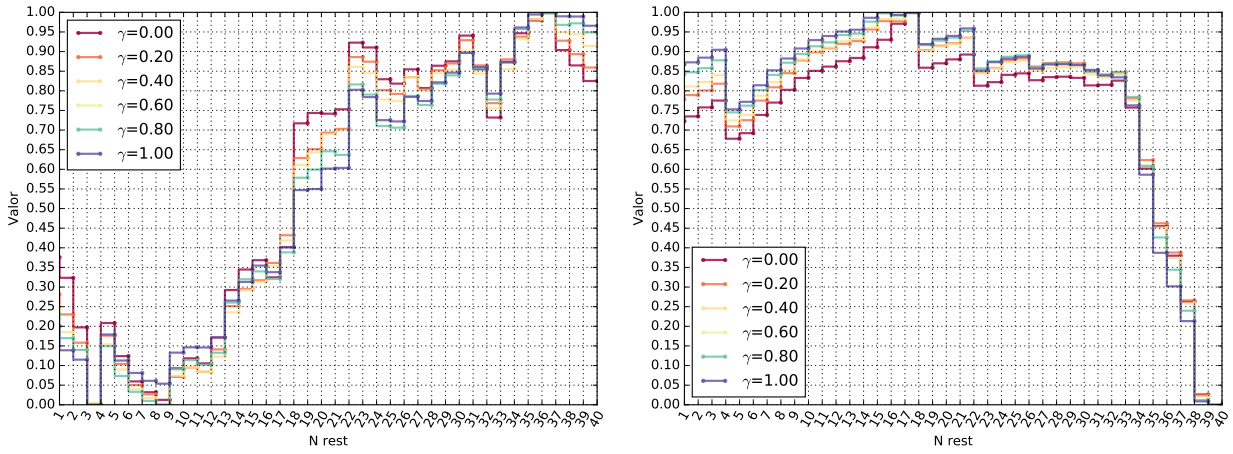


Figura 4.13: Elipse de cada componente de modelo *a posteriori* graficando cada Gaussiana con 0.9 de probabilidad. El mecanismo se ejecuta considerando $r^2 = 6$ y $\gamma = 0.5$.

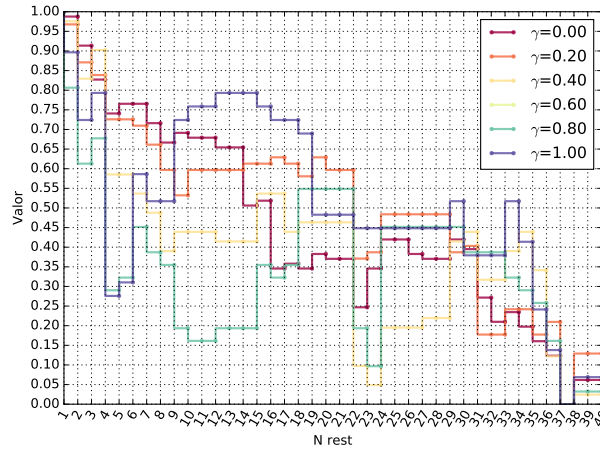
Cabe mencionar que $r^2 \in [1, \infty)$ donde para valor 1 todas las componentes son afectadas por el mecanismo y en ∞ ninguna. Con esto, el problema se traduce a establecer cuál es la cantidad óptima de componentes del total a transformar. De acuerdo al ejemplo anterior se calculan las métricas espaciales *Silhouette*, *Aciertos* y *Desviaciones* como se aprecia en

la Figura 4.14 donde se varía el umbral r^2 que es inversamente proporcional a N_{rest} (que corresponde al número de componentes afectados por la restricción) y la separación de las nuevas componentes con respecto a la central (γ).



(a) *Silhouette*

(b) *Aciertos*



(c) *Desviaciones*

Figura 4.14: Métricas espaciales para modelo *a posteriori* de 40 componentes iniciales, utilizando el mecanismo para diferentes γ y diferentes r^2 tal que N_{rest} (cantidad de componentes afectadas por el mecanismo) cambie uno en uno. Obs: Los criterios están normalizados entre 0 y 1.

La Figura 4.14a indica que a partir de $N_{\text{rest}} = 9$ independiente de la separación de los nuevos componentes se tiene un *clustering* más representativo de los datos con un *peak* en 37 ($40 + 37 \cdot 2 = 114$ componentes totales) y posteriormente se tiene una tendencia de disminución, indicando saturación en la cantidad de componentes generados. Pasando al criterio de Aciertos de la Figura 4.14b las tendencias son muy similares para distintos γ habiendo un *peak* en 18 ($40 + 18 \cdot 2 = 76$ componentes totales) indicando que el mecanismo surte un efecto positivo en la caracterización espacial, pero llegado a un cierto punto existe una clara tendencia de disminución de aciertos como en el caso de $N_{\text{rest}} = 34$. Finalmente, la métrica de Desviaciones de la Figura 4.14c no le es indiferente el valor de γ para los diferentes valores de N_{rest} . Si bien en promedio se puede identificar una tendencia a la baja, cuando se aumenta N_{rest} , no se puede distinguir puntos codos o algún elemento para determinar

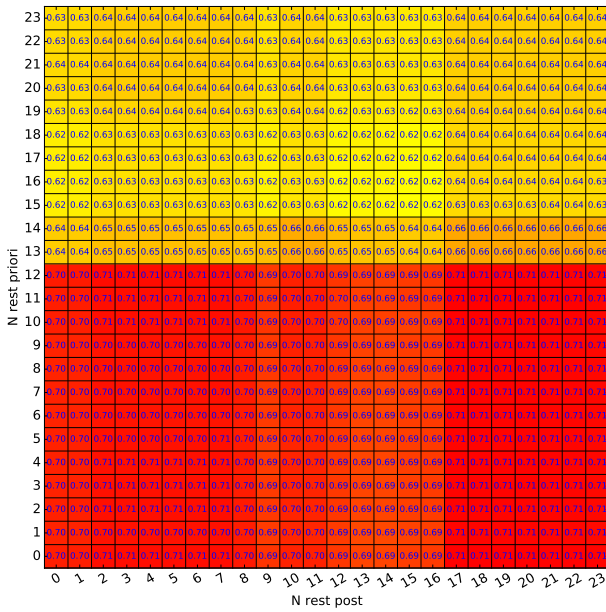
saturación de componentes. El comportamiento de Desviaciones no resulta extraño, debido a que la desviación estándar de las clasificaciones de datos para distintos componentes es muy sensible, especialmente si se agregan componentes espaciales mediante el procedimiento ya descrito.

Hasta ahora se ha analizado solo el comportamiento espacial, es por ello que con el objetivo de determinar el desempeño predictivo se escoge de los estudios anteriores los modelos *a priori* y *a posteriori* con 118 y 33 número de componentes respectivamente. En la Figura 4.15 se tiene los resultados predictivos con AUC de uno a tres pasos y el promedio de éstos para diferentes N_{rest} (cantidad de componentes afectadas por un r^2 fijo), además de considerar $\gamma = 0.5$ puesto que no representa variabilidad en los resultados espaciales significativas para al menos las métricas *Silhouette* y *Aciertos*.

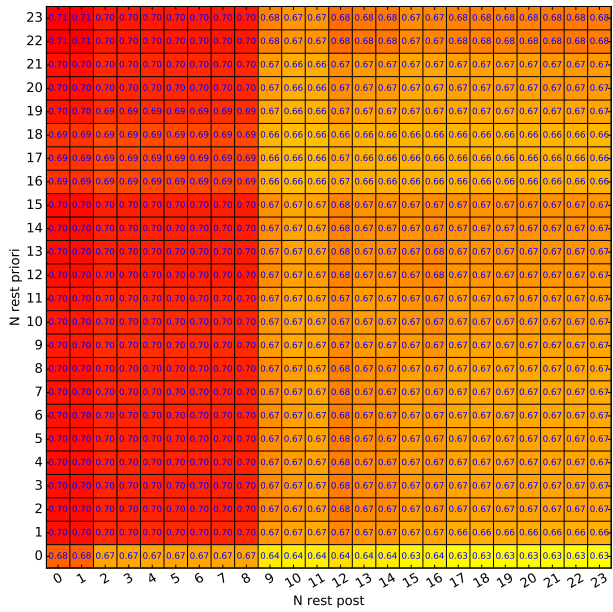
La Figura 4.15a indica una clara división en desempeño al pasar de $N_{\text{rest}}^{\text{priori}}$ 12 a 13 y el máximo se encuentra para $N_{\text{rest}}^{\text{priori}} = 12$ y $N_{\text{rest}}^{\text{post}} = [2, 3, 4, 5, 6]$. Para el caso de la Figura 4.15b ahora la división se obtiene entre $N_{\text{rest}}^{\text{post}} = 8$ y 9 independiente del valor para el caso *a priori*, donde los óptimos predictivos se encuentran al lado izquierdo de esa división. Para el caso de la Figura 4.15c, la división es la misma, pero los óptimos se concentran al lado derecho (opuesto al caso anterior). Por último, para el caso promedio que se aprecia en la Figura 4.15d, los óptimos se concentran en el cuadrante inferior derecho. Es importante notar que $N_{\text{rest}}^{\text{priori}} = 0$ y $N_{\text{rest}}^{\text{post}} = 0$ es el caso base cuando no se modifica ninguna componente por el mecanismo, y como se constata en los distintos pasos de predicción y el promedio, el valor de AUC para el caso base es siempre superado por alguna configuración de $N_{\text{rest}}^{\text{priori}}$ y $N_{\text{rest}}^{\text{post}}$ determinada, pero también es importante notar que no siempre es el valor menor dentro de las matrices, como en el caso de predicción a un paso donde su valor es 0.7 siendo muy cercano a 0.71 que es el máximo.

Discusión

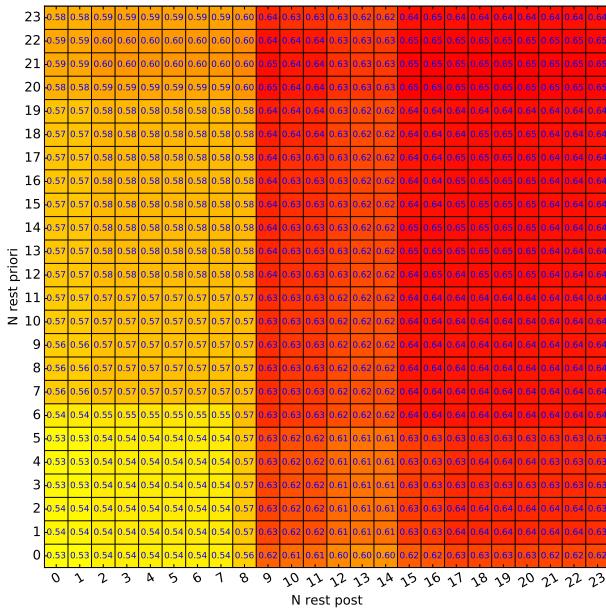
Con este procedimiento ineludiblemente se está aumentando la cantidad de componentes, por lo que la caracterización temporal podría verse dificultada con menos cantidad de eventos criminales por kernels, pese a lo anterior, las series de tiempo estarían representando realmente una dinámica local descorrelacionando la ocurrencia de eventos lejanos geográficamente. Asimismo, nuevamente surge la pregunta: ¿Existe una correlación entre los resultados predictivos con respecto a los resultados de las métricas espaciales mostradas?. No es fácil de terminar esta relación puesto que cada paso de predicción tiene un resultado particular, para lo cual es mejor focalizarse en el caso promedio. Tal como lo indica la Figura 4.15d los óptimos se encuentran aproximadamente en el cuadrante inferior derecho, lo que indica que el modelo *a priori* se satura cuando $N_{\text{rest}}^{\text{priori}}$ supera 12, en cambio el modelo *a posteriori* no se satura, pero requiere de un mínimo de 9 $N_{\text{rest}}^{\text{post}}$. El fenómeno anterior puede deberse a que el *peak* de *Silhouette* del modelo *a priori* se encuentre en 12, sin embargo no se puede inferir directamente por cuanto en cada paso de predicción los delitos de validación son naturalmente diferentes, por tanto la validación debiese hacerse con una gran variedad de fechas y estudiar en promedio la correlación entre comportamiento espacial y predictivo.



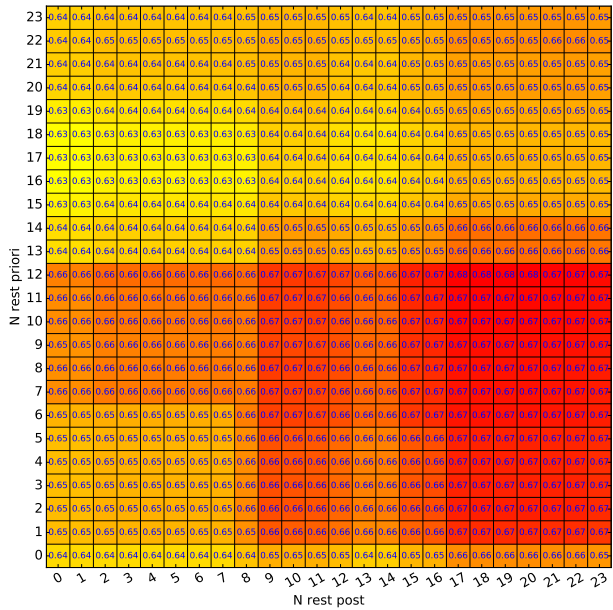
(a) AUC predicción a un paso.



(b) AUC predicción a dos pasos.



(c) AUC predicción a tres pasos.



(d) AUC promedio de los tres pasos.

Figura 4.15: Valores de AUC a un paso y promedio de tres pasos para diferentes componentes espaciales de modelo *a priori* y *a posteriori*.

4.1.7. Estudio de Inclusión de Calles en la Caracterización de Riesgo

Incluir calles como elemento geográfico en la caracterización de riesgo criminal es clave, puesto que condiciona el espacio a segmentos específicos. Puente Alto posee aproximadamente 40.000 segmentos específicos de calles, por lo que situar Gaussianas en cada uno de ellos vuelve el problema intratable desde el punto de vista de costo computacional. Es por ello,

que surge el concepto de calles de importancia, las cuales son aquellas donde mayor riesgo criminal reside. Para calcular ese riesgo criminal, se genera una superficie de riesgo mediante KDE explorando el mejor *bandwidth* (que maximiza la verosimilitud de la caracterización con respecto a los datos) para un número determinado de ellos que en este caso es una partición de 10 valores para el intervalo $[1e-5, 1e-4]$. Luego el valor de riesgo de los segmentos corresponde a la evaluación de su punto medio en la superficie calculada con KDE. Como se aprecia en la Figura 4.16 se ha calculado el riesgo criminal *a priori* (ver Figura 4.16a) y *a posteriori* (ver Figura 4.16b) condicionado a las calles de la comuna de Puente Alto. Para tener una mayor comprensión de los valores de riesgo de los segmentos, se ha graficado en escala de colores (de azul hasta rojo) aquellos segmentos dentro del percentil 80 de riesgo en adelante, y aquellos debajo se mantienen con un tono azul.

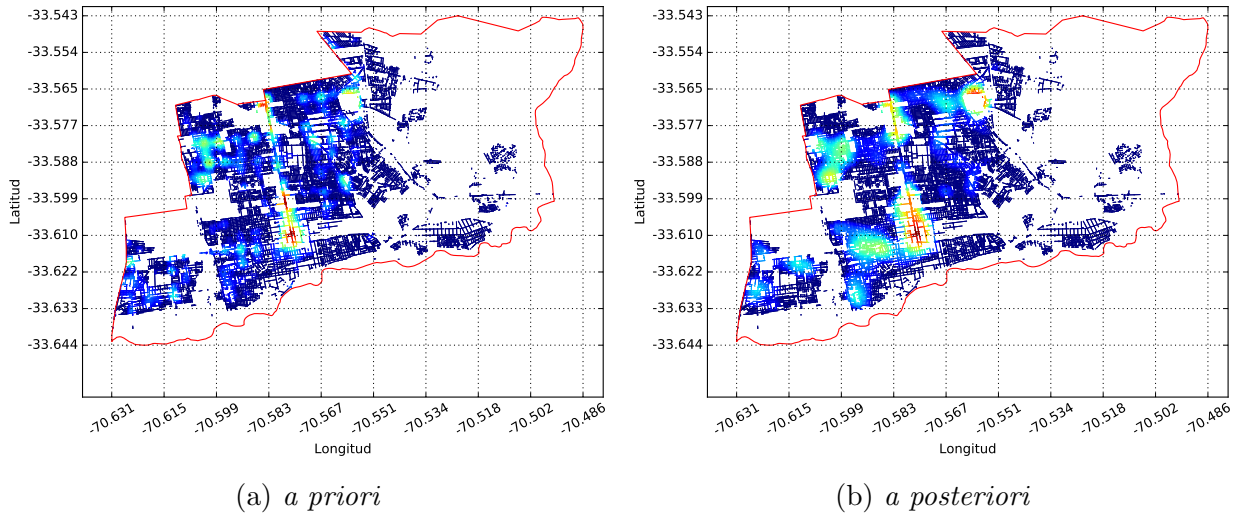


Figura 4.16: Gráfico de calles de la comuna de estudio diferenciando el valor de riesgo de cada una tanto para modelo *a priori* como *a posteriori*. Para este ejemplo las calles de importancia se grafican en escala de colores a partir del percentil 80 de riesgo, y las que están bajo este umbral son pintadas azul.

El proceso de entrenamiento temporal acá difiere del método convencional puesto que los segmentos de calle son numerosos, y por ende la cantidad de eventos asignado a cada segmento sería ínfimo como para generar una serie de tiempo adecuada. Es por ello, que se realiza un modelo GMM de los datos de entrenamiento según corresponda (caso *a priori* como *a posteriori*) y se clasifican los puntos medios de los segmentos de acuerdo a cada componente espacial. Luego la ponderación temporal de cada segmento de calle corresponderá a la serie de tiempo de predicción a la cual pertenece.

Con respecto a las medidas de predicción, para poder generar un HRP y TIP adecuado (puesto que ya no es correcto tan solo proyectar los delitos de predicción en segmentos de calles y ver cuáles se encendieron por el modelo), se genera una superficie de riesgo mediante KDE de los delitos de predicción, y luego se evalúan los puntos medios de los segmentos de calles en esa superficie. Con esto es posible contrastar el riesgo criminal que el modelo predice con respecto al que indica la superficie de riesgo de predicción.

Como se aprecia en la Figura 4.17 los AUC's de predicción se encuentran sobre el valor 0.9, lo que podría indicar un aumento de precisión en el modelo en comparación al uso de grillas.

Sin embargo cabe recalcar que estas métricas adaptadas no son realmente comparables con los análisis de los estudios anteriores, puesto que a diferencia de los anteriores en donde se trabaja con grillas y se verifica la cantidad de aciertos en función de un área determinada, en este caso se verifica la cantidad de riesgo de predicción proyectado en los segmentos de calles predichos por el modelo.

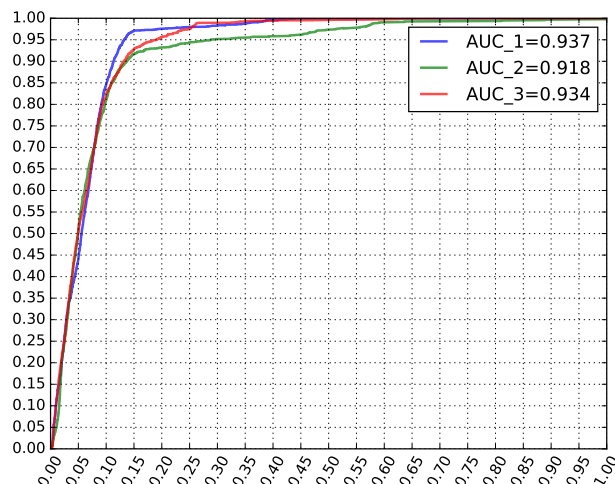


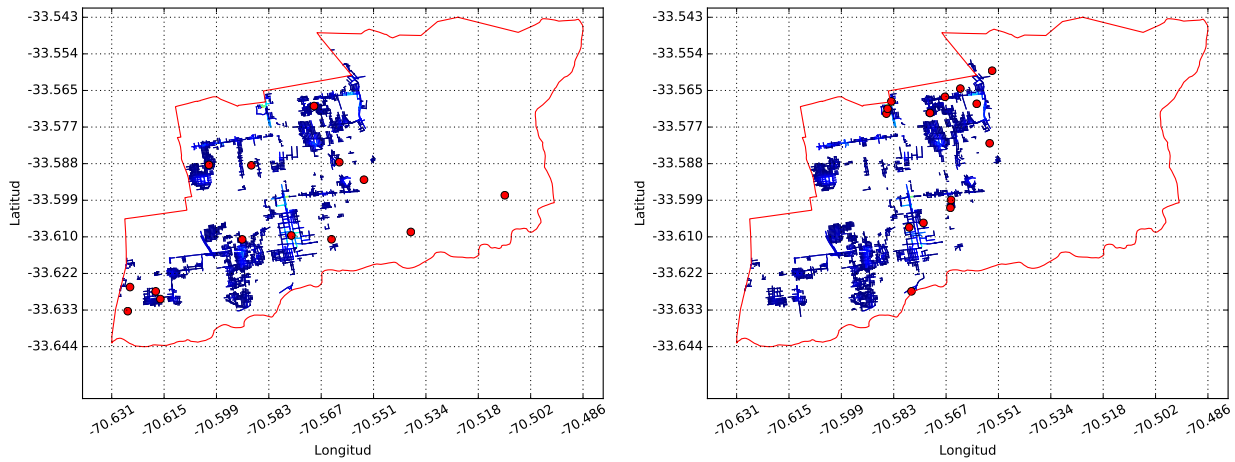
Figura 4.17: Curva HRP vs TIP para modelo con geometría de calles para la fecha 2016-05-03.

En la Figura 4.18 se presenta el resultado para tres pasos de predicción en la fecha indicada, visualizando el 20 % de los segmentos totales de la comuna, donde el TIP del riesgo acumulado está en el rango de 0.93 y 0.7 considerando los tres pasos de predicción.

Discusión

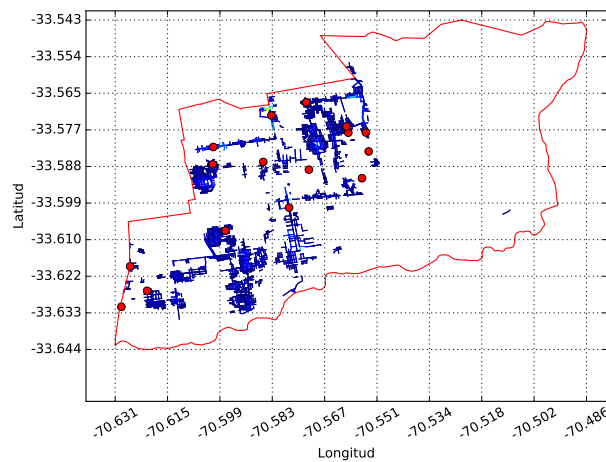
Las predicciones de uno a tres pasos de predicción de las Figuras 4.18a, 4.18b y 4.18c demuestran en comparación al uso de grillas una caracterización del riesgo espacial más específica, lo que en un primer término supondría una ventaja, puede transformarse en una desventaja si las predicciones no son exactas. Si bien, no todos los eventos criminales de predicción (en rojo) son capturados en el 20 % de los segmentos de mayor riesgo de la comuna, es posible detectar patrones espaciales en cada paso de predicción conformado por conjuntos de segmentos, los cuales en la literatura son llamados *line-spots*. Este tipo de caracterizaciones no solo ofrece demarcar las zonas de mayor riesgo, sino que entrega nociones de rutas que podrían trazar patrullas policiales para un barrido eficaz de la comuna.

Volviendo a la implementación del modelo, para tener una idea más acabada de éste, aparte de realizar un estudio extensivo considerando periodos de tiempo prolongados y más zonas de estudio, se debe comparar el desempeño predictivo del algoritmo utilizando variantes: ¿cuál es el desempeño al utilizar predicciones sin influencia de la ponderación temporal (predicciones estáticas) versus al utilizarlas (predicciones dinámicas)?. Si bien en la siguiente sección se remota este punto con el algoritmo convencional, es importante mencionar que el uso de geometrías es útil siempre y cuando los eventos criminales trabajados sean subconjunto o estén proyectadas en las geometrías trabajadas (sin producir un sesgo importante). Además



(a) Predicción a un paso.

(b) Predicción a dos pasos.



(c) Predicción a tres pasos.

Figura 4.18: Resultados de algoritmo con segmentos de calles para tres pasos de predicción para la fecha 2016-05-03 mostrando el 20% de los segmentos de mayor riesgo y en puntos rojos los delitos de predicción de la fecha.

que se debe tener cuidado en las métricas de desempeño a utilizar, que en este caso se adaptaron para esta metodología en particular, pero no resultan ser comparables con los otros estudios. Por último, a modo de conclusión parcial, en la medida que los modelos sean más sofisticados, potencialmente podrían alcanzarse mejores resultados que modelos más simples siempre y cuando las hipótesis, y por ende la cantidad de parámetros extra sean estimables certeramente con la cantidad de datos disponibles.

4.1.8. Resultados Predictivos Generales

Los estudios de las secciones anteriores ilustraron la influencia de los diferentes parámetros para la caracterización del riesgo criminal en términos espaciales y predictivos. En esta sección se realiza un análisis de los resultados predictivos para un periodo de tiempo extenso con el objetivo de validar la metodología. En este caso se realizan tres diferentes tipos de análisis:

En primer lugar, se ejecuta la versión estática del algoritmo, vale decir, sin generar ningún tipo de procesamiento temporal fijando $R_j^{pr}(t) = 1$ y $R_j^{po}(t) = 1$, esto con el objetivo de determinar cuán poderosas son las predicciones solo a nivel espacial y también determinar cuál de los modelos espaciales (*a priori* y *a posteriori*) posee mayor preponderancia en las predicciones variando los parámetros $\lambda = \lambda^{pr}$ y $\lambda^{po} = 1 - \lambda^{pr}$ (ver (3.37)). En segundo lugar se generan predicciones perfectas, vale decir, se construyen las series de tiempo asociadas a cada kernel, de tal manera que se genere un *peak* cada vez que un delito de predicción ocurra, esto con el objetivo de determinar cuál es la capacidad máxima predictiva de la metodología en términos espaciales considerando un perfecto desempeño temporal. Finalmente el tercer análisis consiste en poner a prueba tanto la capacidad predictiva espacial como temporal de la metodología utilizando las series de tiempo de entrenamiento de cada kernel (enfoque estándar). Cabe mencionar que se pone a prueba el algoritmo sin restricciones espaciales, ni utilizando calles.

Análisis Estático

Para este caso la influencia temporal queda descartada, por lo cual es posible analizar la influencia espacial de los modelos *a priori* y *a posteriori* variando el parámetro λ . Como esta versión del algoritmo es estática, se la considera equivalente a la implementación de la **metodología espacial KDE**, en donde se toma una ventana de datos hacia atrás a partir de la fecha consultada (que por lo general son tres meses hacia atrás) y luego las predicciones son el resultado de calcular la superficie de riesgo de los datos con un *bandwidth* que por lo general es de 150[m]. Para este estudio se computan predicciones de 1 a 3 pasos considerando la métrica AUC y valores de $\lambda = [0, 0.05, \dots, 1.0]$, los datos de entrenamiento *a posteriori* son de dos meses previos a la fecha especificada y los datos *a priori* tienen duración de un año. La cantidad de componentes espaciales se fija en 118 y 33 para los modelos espaciales respectivamente y para simplificar cálculos, se fija la cantidad de iteraciones del algoritmo EM en 300. Como se aprecia en la Figura 4.19 se calcula para el primer semestre del 2015 la matriz de promedios de AUC's (Figura 4.19a) y la matriz de cuentas, la que contabiliza las veces que se obtuvieron los AUC's máximos en esas fechas (Figura 4.19b). Como lo indican las matrices, los mejores resultados (mayor AUC) se obtienen cuando el λ se acerca a 1.0, lo cual implica que las mejores caracterizaciones se dan cuando se escoge el modelo *a priori* como preponderante. Esto no resulta extraño puesto que al utilizar una mayor cantidad de datos, necesariamente se está caracterizando mejor el riesgo criminal, salvo que al priorizar tendencias históricas no se están capturando las últimas tendencias del modelo *a posteriori*.

Si bien la tendencia según las dos matrices radica en λ altos, con lo cual se infiere que el primer semestre del 2015 tiende a poseer *Hot-Spots* sin mucha variabilidad, es posible apreciar diferencias entre los mejores λ según la Tabla 4.4, lo cual indica que existen fechas en donde sí hay que considerar como preponderante la caracterización *a posteriori*.

Adicionalmente, se puede estudiar el comportamiento de λ para las fechas 2016-01-01 a 2016-05-01, y establecer si entre 2015 y 2016 hubo una importante modificación de los mejores λ , ya que de no haber cambios importantes, se podría utilizar como entrada los mejores λ de cada día del 2015 con su homólogo del 2016. Como se aprecia en la Figura 4.20 y en el resumen de la Tabla 4.5, los λ óptimos del 2016 están más desplazados al valor uno con

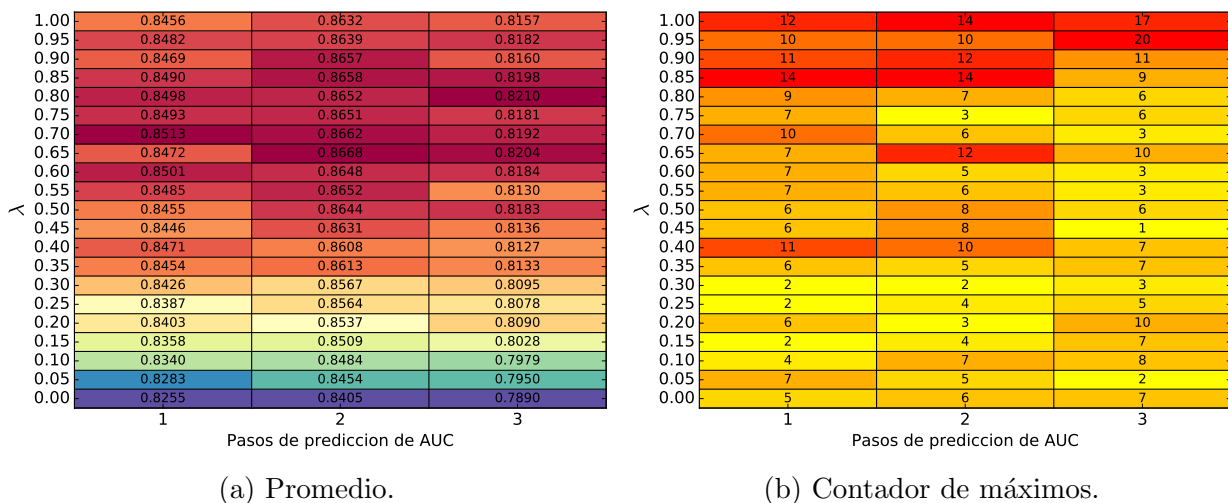


Figura 4.19: Valor promedio de AUC de un 1 a 3 pasos y cantidad de veces que AUC fue máximo para diferentes valores de λ para las fechas 2015-01-01 a 2015-05-01.

	Promedio	Máximos
AUC 1	0.70	0.85
AUC 2	0.65	0.85 y 1.00
AUC 3	0.80	0.95

Tabla 4.4: Resumen mejores λ análisis estático fechas 2015-01-01 a 2015-05-01.

respecto al año anterior, pero de todas maneras no son variaciones significativas en promedio, por lo cual habría que analizar qué ocurre en cada fecha específica.

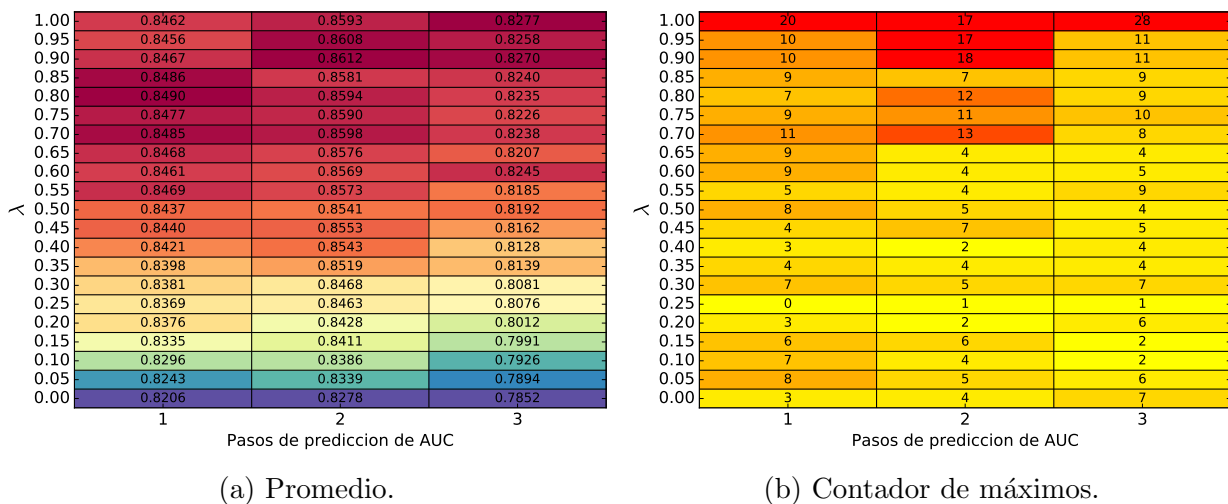


Figura 4.20: Valor promedio de AUC de un 1 a 3 pasos y cantidad de veces que AUC fue máximo para diferentes valores de λ para las fechas 2016-01-01 a 2016-05-01.

Es importante mencionar que los AUC's de predicción para este análisis estático se enmarcan en rangos definidos como se aprecia en la Tabla 4.6, los cuales no llegan a 0.9, pero tampoco bajan de 0.78. Estos rangos se comparan con la versión **estándar** y **series de tiempo perfectas** del algoritmo que se explican a continuación.

	Promedio	Máximos
AUC 1	0.80	1.00
AUC 2	0.90	0.90
AUC 3	1.00	1.00

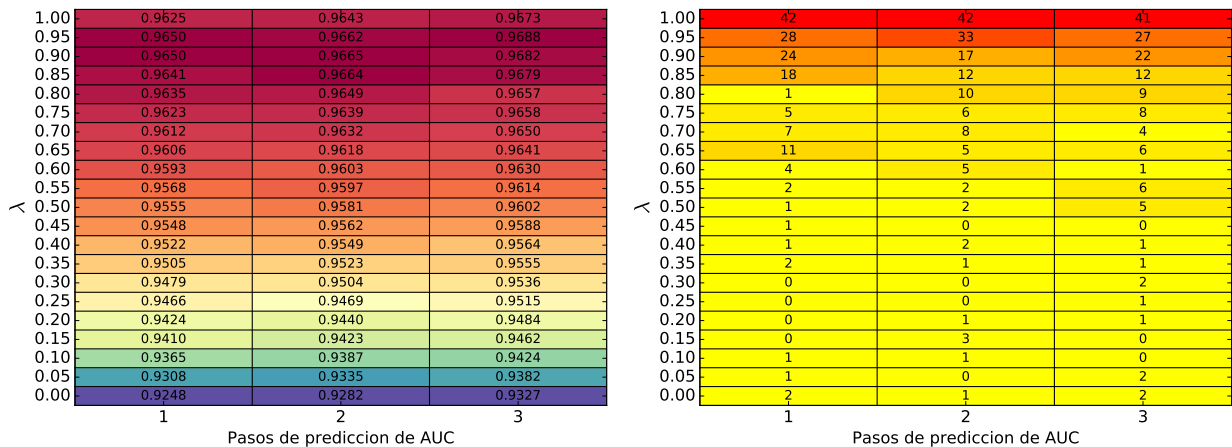
Tabla 4.5: Resumen mejores λ análisis estático fechas 2016-01-01 a 2016-05-01.

	rango año 2015	rango año 2016
AUC 1	[0.8513 – 0.8215]	[0.8490 – 0.8206]
AUC 2	[0.8668 – 0.8405]	[0.8612 – 0.8278]
AUC 3	[0.8210 – 0.7890]	[0.8277 – 0.7852]

Tabla 4.6: Rangos de AUC de predicción para años 2015 y 2016 del análisis estático

Análisis con Series de Tiempo Perfectas

En este análisis el objetivo es determinar una cota máxima de los AUC's de predicción, para ello se realiza el entrenamiento espacial ya explicado anteriormente con la diferencia que para el entrenamiento temporal las series de tiempo se construyen de tal manera que calcen perfectamente con los delitos de validación. Si este procedimiento se hiciera también para el caso espacial, se esperarían AUC's con valor 1.0, es por ello que solo realizándolo para las series temporales se analiza cuán imperfectas son las caracterizaciones espaciales. Como se aprecia en la Figura 4.21 computando la matriz de promedios (Figura 4.21a) y el contador de máximos (Figura 4.21b), los λ máximos no son menores 0.9 como se aprecia en la Tabla 4.7, superando ampliamente los valores del estudio anterior.



(a) Promedio.

(b) Contador de máximos.

Figura 4.21: Valor promedio de AUC de un 1 a 3 pasos y cantidad de veces que AUC fue máximo para diferentes valores de λ para las fechas 2015-01-01 a 2015-05-01.

Asimismo, se presentan los resultados del segundo semestre del año 2016 dados por la Figura 4.22 y la Tabla 4.8 donde los λ son similares al primer semestre del 2015, sin embargo los rangos de AUC's como se aprecia en la Tabla 4.9 son levemente mejores para el periodo del 2016 con respecto al año anterior.

	Promedio	Máximos
AUC 1	0.95 y 0.90	1.00
AUC 2	0.90	1.00
AUC 3	0.95	1.00

Tabla 4.7: Resumen mejores λ análisis series de tiempo perfectas fechas 2015-01-01 a 2015-05-01.

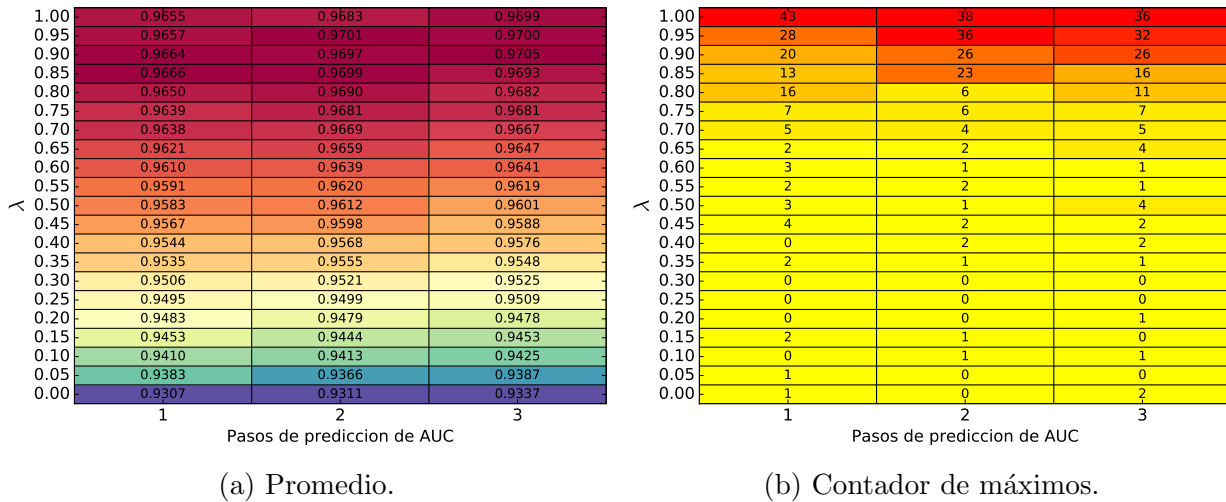


Figura 4.22: Valor promedio de AUC de un 1 a 3 pasos y cantidad de veces que AUC fue máximo para diferentes valores de λ para las fechas 2016-01-01 a 2016-05-01.

	Promedio	Máximos
AUC 1	0.85	1.00
AUC 2	0.85	1.00
AUC 3	0.90	1.00

Tabla 4.8: Resumen mejores λ análisis series de tiempo perfectas fechas 2016-01-01 a 2016-05-01.

Finalmente, cabe mencionar que los rangos de AUC's para este estudio considerando los dos periodos se encuentran en el intervalo $[0.93,0.97]$ a diferencia del estudio anterior que es de $[0.78,0.86]$ aproximadamente.

	rango año 2015	rango año 2016
AUC 1	$[0.9650 - 0.9248]$	$[0.9666 - 0.9307]$
AUC 2	$[0.9665 - 0.9282]$	$[0.9699 - 0.9311]$
AUC 3	$[0.9688 - 0.9327]$	$[0.9705 - 0.9337]$

Tabla 4.9: Rangos de AUC de predicción para años 2015 y 2016 del análisis con series de tiempo perfectas

Análisis Predicciones Estándar

En este análisis, se itera el algoritmo generando las caracterizaciones espaciales como temporales, considerando para el entrenamiento espacial $N^{\text{pri}} = 118$, $N^{\text{po}} = 33$ y $n_{iter} = 300$ fijos en pos de no tener un costo computacional elevado, y en cuanto al entrenamiento temporal, se realiza una reconstrucción mediante la transformada Wavelet de las series de entrenamiento para realizar predicción de ellas como se indica en la Sección 3.4.1. Los resultados, se aprecian en la Figura 4.23 donde los valores de λ tanto para el caso de la Figura 4.23a y Figura 4.23b son variados, pese a que la tendencia esté por sobre el valor 0.5. Como puede apreciarse en la matriz de cuentas, los pasos de predicción se diferencian notoriamente, habiendo acumulación de valores λ en 0.45, 0.2, 0.1, e incluso 0.0.

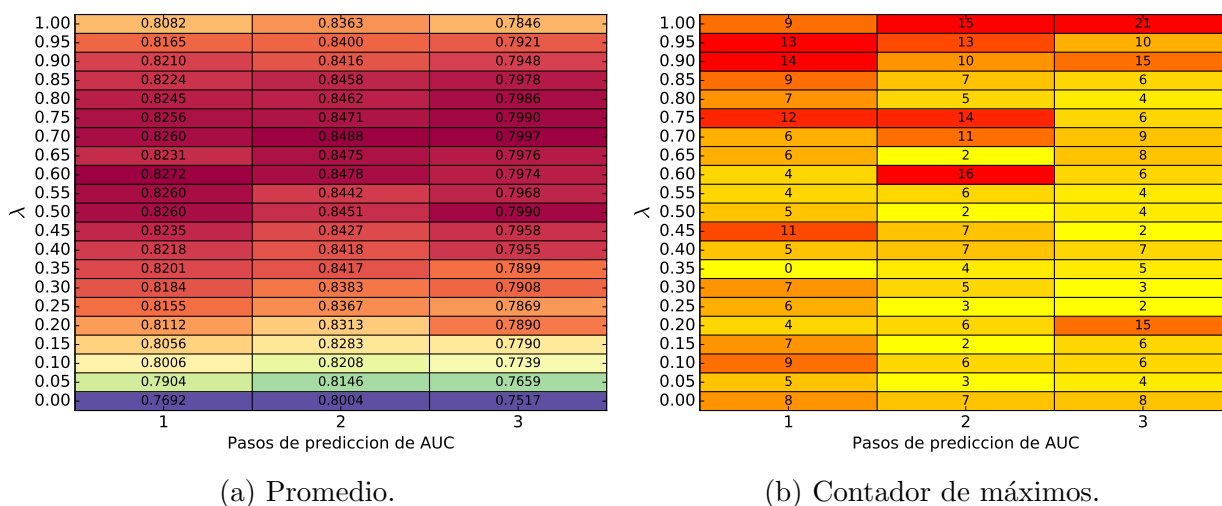


Figura 4.23: Valor promedio de AUC de un 1 a 3 pasos y cantidad de veces que AUC fue máximo para diferentes valores de λ para las fechas 2015-01-01 a 2015-05-01.

	Promedio	Máximos
AUC 1	0.60	0.90
AUC 2	0.70	0.60
AUC 3	0.70	1.00

Tabla 4.10: Resumen mejores λ análisis predicciones estándar fechas 2015-01-01 a 2015-05-01.

Estas variaciones se deben interpretar como la manera en que los desajustes del entrenamiento temporal y espacial son suplidos considerando las últimas tendencias. Este fenómeno no es ajeno al primer semestre del 2016 tampoco, aunque no es tan notorio como el 2015 (ver Figura 4.24a y Figura 4.24b).

Como se verifica en la Tabla 4.11 los mejores λ según la matriz de promedios para los primeros pasos de predicción se sitúan en 0.6 aproximadamente mientras que en el tercero en 0.85, a diferencia de la matriz de cuentas que entrega λ muy cercanos a 1.0.

Resumiendo los rangos de AUC's promedios para los periodos como se aprecia en la Tabla 4.12 no son significativamente diferentes entre sí estando aproximadamente en el intervalo

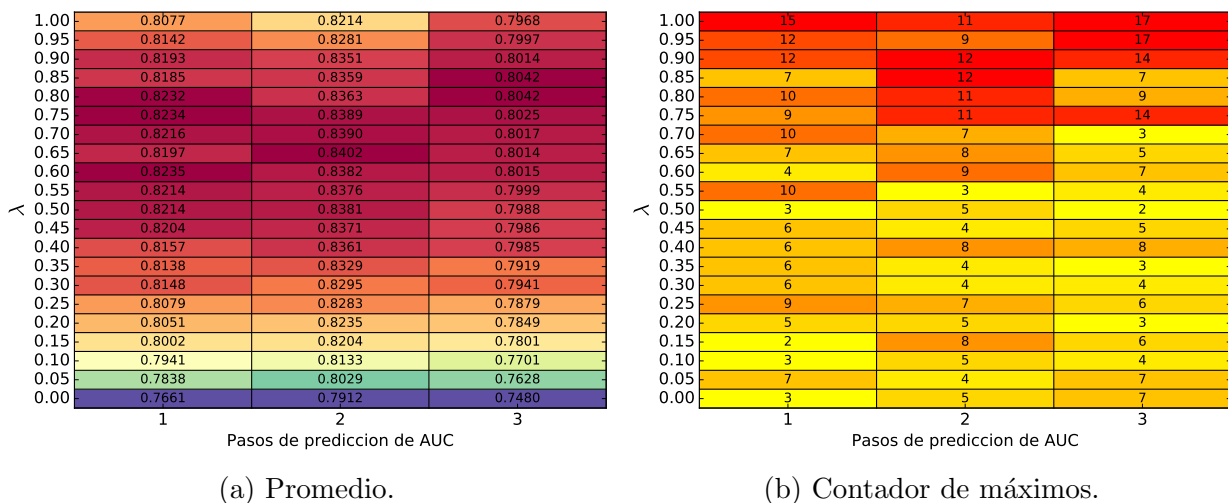


Figura 4.24: Valor promedio de AUC de un 1 a 3 pasos y cantidad de veces que AUC fue máximo para diferentes valores de λ para las fechas 2016-01-01 a 2016-05-01.

	Promedio	Máximos
AUC 1	0.60	1.00
AUC 2	0.65	0.90 y 0.85
AUC 3	0.80 y 0.85	1.00 y 0.95

Tabla 4.11: Resumen mejores λ análisis series de tiempo estándar fechas 2016-01-01 a 2016-05-01.

[0.74,0.84]. Contrastando con los análisis anteriores, en el caso de **series perfectas** se tiene [0.93,0.97] y en el **estático** [0.78,0.86].

	rango año 2015	rango año 2016
AUC 1	[0.8272 – 0.7692]	[0.8235 – 0.7661]
AUC 2	[0.8488 – 0.8004]	[0.8402 – 0.7912]
AUC 3	[0.7997 – 0.7515]	[0.8042 – 0.7480]

Tabla 4.12: Rangos de AUC de predicción para años 2015 y 2016 del análisis con series de tiempo estándar

Discusión

Es claro que el algoritmo con series temporales perfectas tenga un rango de AUC superior a los otros, sin embargo queda abierta la pregunta con respecto a la versión estática y estándar: ¿Por qué para ciertas fechas un modelo estático tiene mejor desempeño predictivo que uno dinámico?, ¿Es un comportamiento específico a los periodos de tiempo analizados o a la zona de estudio?. Se debe considerar que los datos tanto de entrenamiento como de validación están condicionados al actuar de Carabineros de Chile en los periodos estudiados. Lo anterior implica que si el algoritmo estático da mejor que el dinámico en las fechas estudiadas no necesariamente entrega información concluyente para determinar qué versión del algoritmo

se deba utilizar en el futuro. Esto porque en el largo plazo contar con modelos estáticos puede mermar la efectividad de una logística de patrullaje “estático”. Si se revisa todos los días los mismos sectores, siguiendo las mismas superficies de riesgo estáticas, el patrullaje se vuelve predecible, y por lo tanto “evitable”. Con ello, la única manera de medir realmente el algoritmo estándar (dinámico) es mediante una marcha blanca para constatar cómo se adapta a los cambios de patrones criminales. De todas maneras en la siguiente sección se comprueba en qué comunas funciona mejor el enfoque estático, y en cuales el dinámico⁶.

4.2. Resultados de Metodología para Diferentes Comunas

En este estudio, con el objetivo de generar resultados para un mayor espectro geográfico, se adicionan más comunas de la Región Metropolitana como se aprecia en la Figura 4.25, las cuales son: Macúl, Ñuñoa, Peñalolén, Providencia, Colina, Conchalí, Huechuraba, Independencia, Lampa, Quilicura, Recoleta, Tiltill y Maipú.

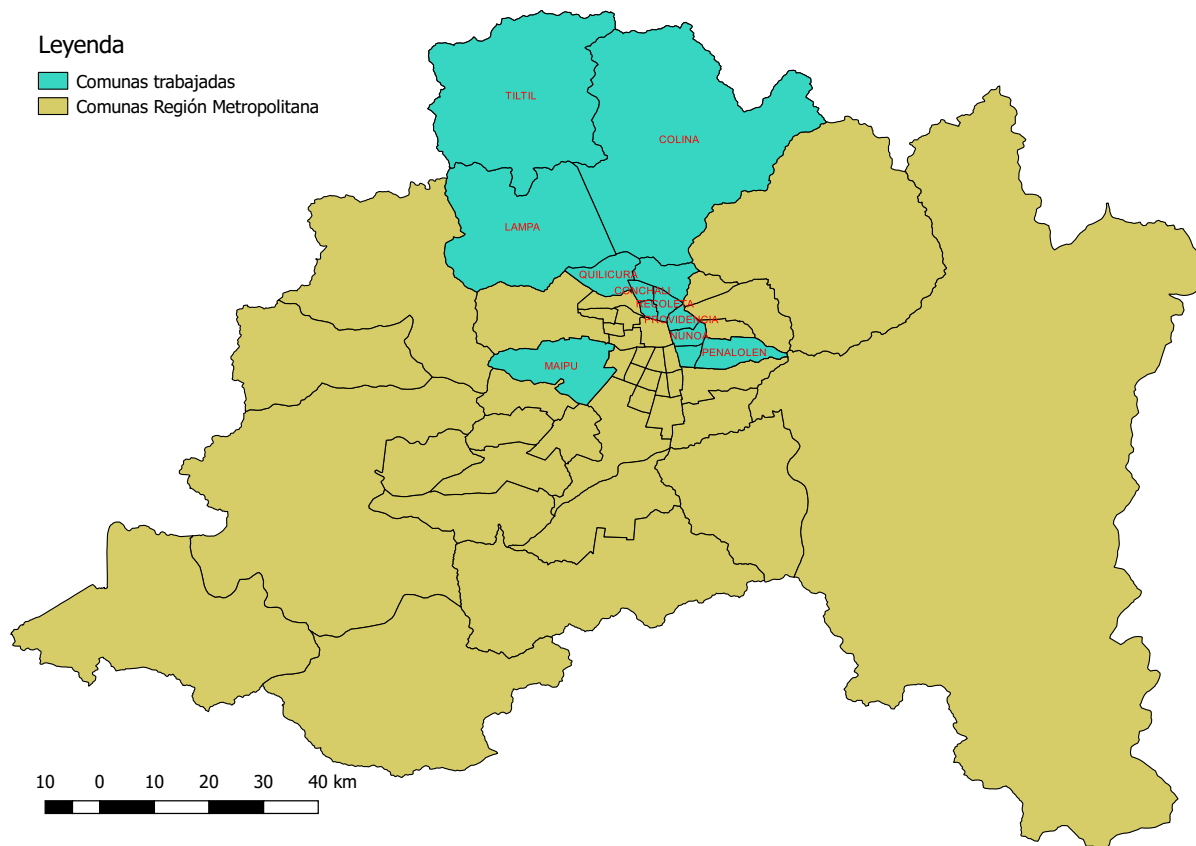


Figura 4.25: Comunas de estudio de la Región Metropolitana.

⁶Independiente del fenómeno ya explicado, los resultados para diferentes comunas puede dar también luces del dinamismo de los patrones criminales.

4.2.1. Resultados AUC Promedio para Cada Comuna de Estudio

Este estudio se realiza para el intervalo de fechas contemplado entre 2016-05-01 y 2016-08-01. En cada comuna se ejecutan las tres versiones del algoritmo explicadas en la sección anterior, donde la caracterización espacial es la misma con las siguientes especificaciones:

- La cantidad de componentes del modelo *a priori* se fija como el 2% de la cantidad de eventos criminales *a priori* y en el caso *a posteriori* como el 5% de la cantidad de datos *a posteriori*.
- La cantidad de iteraciones se fijan en 300 del algoritmo EM.
- La ponderación inter-procedimientos λ se fija en 0.9.
- No se utiliza el algoritmo de restricciones espaciales.

Los resultados específicos de las pruebas se encuentran en Anexos, donde se grafican las curvas HRP v/s TIP promedio de cada paso de predicción para las comunas estudiadas (ver Figura 5.2, Figura 5.3, Figura 5.4 y Figura 5.5), donde cabe destacar lo siguiente:

- Las curvas HRP/TIP de los pasos de predicción presentados para la versión con series de tiempo perfectas del algoritmo constituyen una cota superior para las comunas estudiadas.
- Existen comunas donde la versión estática del algoritmo supera en promedio para los tres pasos de predicción y en la mayoría de áreas de cobertura (HRP) al algoritmo dinámico como en la comuna de Ñuñoa.
- Sin embargo existen casos como Maipú donde el algoritmo dinámico supera al estático en la mayoría de las áreas de cobertura.
- En otras comunas, solo los dos primeros pasos de predicción dinámicos son superiores al estático en la mayoría de las áreas de cobertura como es el caso de Macúl.
- Cabe destacar otros casos donde no es posible establecer diferencias significativas entre curvas como el caso de Huechuraba, Colina y Lampa para las áreas de cobertura de hasta el 20%.

De esta manera, en la Figura 4.26 se muestra de manera resumida los AUC's promedios por cada paso de predicción estudiados con las siguientes observaciones:

- Para el primer paso de predicción, en 4 comunas (Macúl, Peñalolén, Huechuraba y Maipú) la versión dinámica del algoritmo supera a la versión estática.
- Para el segundo paso de predicción, en 5 comunas (Macúl, Peñalolén, Providencia, Huechuraba y Maipú) la versión dinámica del algoritmo supera a la versión estática.
- Para el tercer paso de predicción, en 2 comunas (Peñalolén y Maipú) la versión dinámica del algoritmo supera a la versión estática, sin embargo en el caso de Ñuñoa son equiparables.

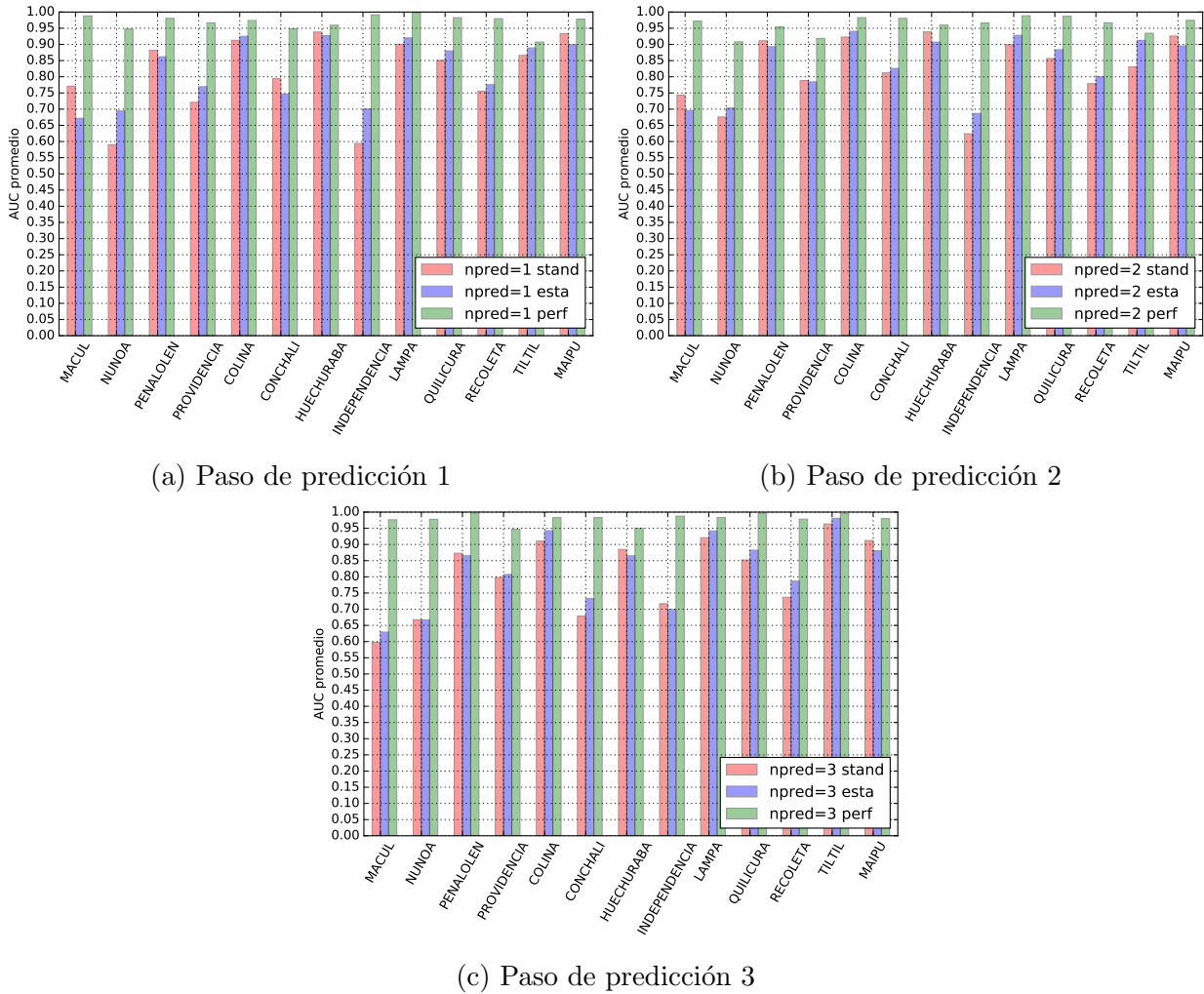


Figura 4.26: AUC's promedio de 1 a 3 pasos de predicción considerando versión **estándar**, **perfecta** y **estática** del algoritmo en el intervalo de fechas 2016-05-01 2016-08-01 para comunas de estudio.

4.2.2. Discusión

- El objetivo de este análisis no es solo validar el algoritmo para diferentes comunas, que para efectos del proyecto en conjunto con Carabineros de Chile es obtener sobre un 35 % de aciertos (TIP) con un 10 % de área de cobertura (HRP) (que por cierto para todas las comunas estudiadas, en promedio se alcanza escogiendo ya sea el algoritmo estático del dinámico cuando corresponda⁷), sino que es importante establecer un criterio de decisión de cuándo escoger la versión estática de la dinámica en cada comuna.
- La problemática anterior está directamente relacionada con cuán estáticos/dinámicos son los *Hot-Spots* de una determinada comuna, vale decir, si a lo largo de los datos históricos incorporados en el entrenamiento de los modelos ha existido movimiento de estos puntos lo suficientemente significativos como para escoger el algoritmo dinámico,

⁷En Ñuñoa se tiene un rango del 20 % al 25 % promedio, por lo cual debería analizarse en más detalle si el bajo porcentaje de aciertos es por la baja cantidad de delitos de entrenamiento/validación o porque los patrones difieren mucho entre lo entrenado versus lo validado.

o el caso contrario como para escoger el estático.

- Lo importante a destacar, es que en la medida que los *Hot-Spots* sean más estáticos, se hace más menos necesario computar los parámetros para construir las ponderaciones temporales, esto porque al calcularlos (que no solo significa un mayor costo computacional) se puede introducir más ruido de predicción de lo usual por medio del cálculo de componentes de frecuencias y fases con incertidumbre (cuando solo era necesario un análisis estático). En caso contrario, queda claro que solo proyectar un mapa estático ante patrones altamente dinámicos no tiene sentido, por lo que entrenar los modelos temporales significaría una ventaja.
- De esta manera queda abierta la posibilidad de introducir criterios que caractericen la dinámica de una comuna para tomar decisiones sobre qué versión del algoritmo es más adecuada utilizar.

4.3. Comparación de Resultados Predictivos con Otros Métodos

En esta sección se compara los resultados obtenidos por el algoritmo propuesto por esta tesis con algunos de los principales algoritmos de la literatura científica. Esta comparación resulta de vital importancia para situar adecuadamente el algoritmo propuesto de acuerdo a las métricas predictivas como la cantidad de eventos criminales acertados para una cantidad de área encendida (que corresponde a evaluar el HRP y TIP de manera puntual) y el AUC. Aunque estas medidas son adecuadas para comparar algoritmos, de todas maneras se deben tomar en cuenta los siguientes factores:

- Tipo de delito utilizado en la predicción.
- Resolución espacial utilizada y características de la zona de estudio.
- Resolución temporal utilizada e intervalo de tiempo de validación predictiva.

Como se aprecia en la Figura 4.27, se presenta la curva HRP/TIP del trabajo de Mohler *et al.* [10] para robos residenciales promedios del año 2015 proveídos por la Policía de los Angeles. Los resultados demuestran que el algoritmo de procesos puntuales auto-excitados resulta ser para todas las áreas pintadas mejor algoritmo predictivo que el modelo prospectivo. Realizando una comparación directa con el algoritmo desarrollado en esta tesis, para un 10% del área cubierta, procesos puntuales auto-excitados solo logra predecir un 25% de los eventos criminales validados, mientras que el algoritmo propuesto en promedio para las comunas trabajadas logra un 65% de aciertos promedio (ver Tabla 4.13). Es importante mencionar que en el algoritmo propuesto de esta tesis se trabaja con delitos que ocurren en espacio de vía pública, a diferencia del trabajo de Mohler *et al.* [10] donde se trabaja con delitos residenciales, lo cual sugiere que la naturaleza del tipo de delito puede distorsionar las comparaciones. Pero además se puede estar distorsionando la comparación al no normalizarse el área total espacial por el área donde nunca ha ocurrido delitos⁸

⁸Lamentablemente no se muestra en el trabajo de Mohler *et al.* una imagen de la zona con todos los delitos georeferenciados para comparar.

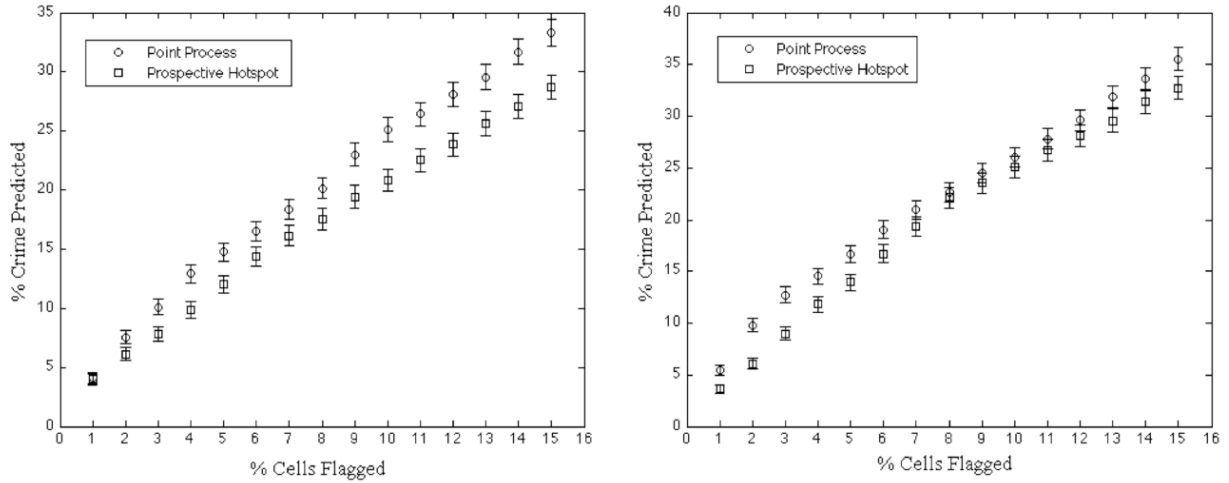


Figura 4.27: Resultados predictivos diarios usando celdas de $200 \times 200 [m^2]$. En cuanto al método prospectivo los parámetros de kernels son 400 [m] espaciales y 8 [semanas] (figura de la izquierda) y parámetros óptimos (figura de la derecha) donde se considera 200 [m] espaciales y 39 [semanas] temporales. La intensidad de fondo para el método de procesos puntuales corresponde a un ancho de banda de 300 [metros] (figura de la izquierda) seleccionado mediante validación cruzada y 130 [m] (figura de la derecha) seleccionado para optimizar el número de eventos criminales a predecir. Fuente: Trabajo de Mohler *et al.* [10]

Por otro lado, también se presenta los resultados del trabajo de Wang *et al.* [4], en donde compara una batería de algoritmos en su estudio como se aprecia en la Figura 4.28 que corresponde al modelo de Hot-Spots, *Generalized Linear Models* (GLM) y modelos aditivos que propone en su trabajo, que básicamente corresponden a modelos logísticos en cada grilla espacial tomando variables de entrada socio-económicas y demográficas. Dentro de sus propias comparaciones, los algoritmos ST-GAM y LST-GAM superan ampliamente al modelo de Hot-Spots y GLM para los meses trabajados, y a modo de ejemplo, para el caso particular de Julio, nuevamente los algoritmos ya nombrados superan para cualquier área de cobertura (HRP) a Hot-Spots y GLM. Realizando la comparación directa con el algoritmo presentado en esta tesis, ST-GAM y LST-GAM presentan para un 10% de cobertura un 60% de aciertos en promedio, mientras que para el algoritmo propuesto de esta tesis se tiene un 65% de aciertos en promedio para las zonas de estudio trabajadas. Estos resultados sugieren que los algoritmos aditivos con el propuesto son similares en términos predictivos, sin embargo es importante recalcar las diferencias en las metodologías: en términos espaciales Wang *et al.* utilizan grillas de $32 \times 32 [m^2]$ a diferencia del algoritmo de esta tesis que es de $150 \times 150 [m^2]$, en cuanto a resolución temporal se ocupan intervalos de 1 mes para 46 meses de base de datos, en cambio en el algoritmo de esta tesis se ocupa una resolución de 8 horas (duración de cada turno policial) considerando una base de datos móvil de 1 año y 2 meses para 5 años de base de datos en total. También en cuanto a los tipos de delitos, en el trabajo de Wang *et al.* se caracterizan y predicen delitos de “allanamiento de morada”, los que en términos de la Tabla 5.1 pueden catalogarse como delitos del agrupado de Robo Violencia, código 809 y 810, siendo un subconjunto de los delitos ocupados en esta tesis correspondiente a los agrupados de Robo Violencia y Robo Fuerza completos. Debido a la naturaleza de los delitos de “allanamiento de morada” es que se hace necesaria una caracterización socio-económica y demográfica de las grillas estudiadas, para establecer los posibles lugares fijos en donde

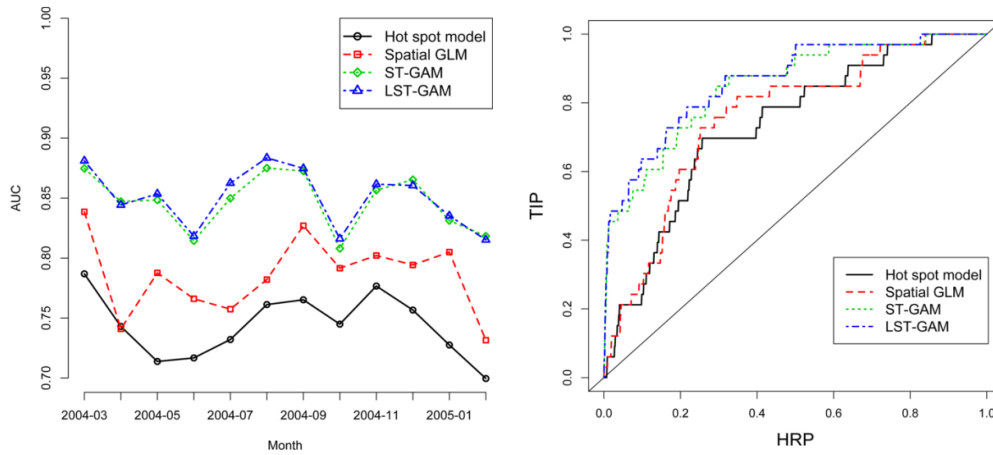


Figura 4.28: AUC para eventos criminales de Charlottesville desde Marzo 2004 a Febrero 2005 (izquierda). Curva HRP/TIP para mes de Julio 2004 (derecha). En estos resultados se compara los algoritmos propuestos por Wang *et al.* consistentes en modelos aditivos GAM (línea verde y azul) con el típico modelo mediante Hot-spots (línea negra) y otro llamado *Generalized Linear Models* (GLM) (línea roja). Fuente: Trabajo de Wang *et al.* [4]

puedan ocurrir mayor cantidad de eventos criminales. Lo anterior contrasta con los tipos de delitos trabajados en esta tesis, puesto que incluyen los que ocurren en vía pública habiendo mayor dificultad en caracterizarlos y predecirlos, debido a que no se ciñen necesariamente por el principio de victimización cercana en espacio y repetida en el tiempo, fenómeno que sí ocurre con delitos residenciales.

Por último, a raíz de las comparaciones anteriores se puede concluir que no existe un modelo universal de riesgo criminal que procese todo tipo de delitos, por ende, cada modelo analizado de la literatura tiene sus propias hipótesis a verificar, donde por supuesto se incluye el tipo de delito a trabajar: en el caso de esta tesis se trabaja con delitos del tipo robo con violencia y fuerza en espacio de vía pública. Pero existen otros estudios que abarcan modelos basados en servicios para predecir hurtos, o modelos que se basan en el fenómeno de victimización repetida en espacio-tiempo para estudiar robos en espacio habitado. Comparar distintos algoritmos ejecutado en diferentes áreas de estudios puede distorsionar las medidas HRP, TIP y AUC, esto porque dado que estas medidas toman en cuenta el total del área de estudio pueden haber zonas con áreas extensas donde no ocurran delitos (desiertos, zonas rurales, entre otros), por lo que un 10 % de un área de cobertura puede representar fácilmente un 100 % de aciertos ya que se pinta el área urbana, mientras que en zonas de estudio donde no existen estas áreas sin delitos, un 10 % de un área de cobertura puede fácilmente representar un bajo porcentaje de aciertos predictivos. En ese sentido, queda generar un *framework* donde se tomen en cuenta estos aspectos para realizar una comparación justa entre algoritmos.

Comuna	% aciertos con 10 % de cobertura
Colina	90
Conchalí	55
Huechuraba	85
Independencia	40
Lampa	85
Macúl	45
Maipú	80
Ñuñoa	30
Peñalolén	75
Providencia	50
Quilicura	65
Recoleta	50
Tiltil	100
Promedio	65.38 %

Tabla 4.13: Porcentaje de aciertos para un 10 % de cobertura para las diferentes comunas trabajadas considerando el mejor resultado según el turno policial y versión del algoritmo (estático o dinámico). Basado en los resultados de Anexos: Resultados AUC promedio para cada comuna de estudio.

Capítulo 5

Conclusiones

En función de los objetivos generales y específicos se concluye lo siguiente:

- Se desarrolla una metodología capaz de caracterizar patrones espacio-temporales de eventos criminales de manera adecuada para la mayoría de las comunas estudiadas de la Región Metropolitana¹. Los tipos de eventos criminales son acotados a DMCS específicamente a los agrupados de Robo con Fuerza y Violencia comprobándose que es posible caracterizarlos mediante una función de intensidad espacio-temporal. Notar que predicciones que se hagan considerando delitos de otra naturaleza (como hurto por ejemplo), no van a ser necesariamente precisas puesto que cada tipo de delito tiene su propio patrón (para el caso del hurto, debiera adicionarse al modelo espacial componentes centrados en servicios, puesto que ese tipo de delito tiende a concentrarse en éstos).
- Se diseña efectivamente una metodología que pueda incorporar tendencias históricas y recientes, analizando en detalle cuáles son las ponderaciones óptimas predictivas en determinados escenarios (estudio del factor *inter-procedimiento*). Se comprueba que por lo general las últimas tendencias de los dos últimos meses son más preponderantes que las tendencias históricas de hace un año.
- Se diseña una metodología predictiva que incorpora tanto las caracterizaciones espaciales como temporales del riesgo criminal. Estas predicciones son completamente dinámicas puesto que cada componente es ponderado por un factor temporal propio y un factor que da cuenta sobre tendencias históricas y recientes. Desde el punto de vista de las métricas de evaluación, las curvas HRP v/s TIP, y el AUC resultan fundamentales para escoger parámetros óptimos en los diferentes estudios espaciales realizados.
- Mediante los estudios realizados, se hallan relaciones interesantes entre los distintos parámetros de la caracterización espacial y el desempeño predictivo (considerando una caracterización temporal fija dada por reconstrucción Wavelet más proyección de señales pretéritas). En ese sentido este trabajo de tesis debe enmarcarse como un primer paso para entender la relación entre parámetros espaciales, y cuáles deben ser sus rangos para cumplir con los aciertos predictivos propuestos. En ese sentido se concluye lo siguiente:

¹en conjunto con Carabineros de Chile se fija un 35% de aciertos predictivos en promedio para un 10% del área de cobertura para las comunas estudiadas.

- En cuanto al estudio de criterios de término de iteraciones espaciales, el criterio de **distancias** resulta clave para determinar cuándo las medias, covarianzas y pesos de una determinada GMM se estabilizan, no significando un importante costo computacional al calcularse. Además, calculando adecuadamente la cantidad de iteraciones del algoritmo EM, se evita la distorsión de las métricas espaciales del estudio de cantidad de componentes iniciales.
 - En cuanto al estudio de cantidad de componentes iniciales, si bien no es muy clara la correlación entre la optimización multi-objetivo y el desempeño predictivo, de todas maneras resultan ser funciones que dan al usuario un mayor entendimiento de qué cantidad de componentes es óptima teniendo en cuenta estos criterios. Puesto que puede fijar las ponderaciones entre los criterios de **Aciertos**, **Desviaciones** o **Silhouette** y así establecer la cantidad de componentes que arrojan.
 - En cuanto al estudio de mecanismo sobre valores propios de matrices de covarianza, se establece que el parámetro fundamental y con una alta incidencia en el desempeño predictivo está dado por la cantidad de componentes a iterar con el mecanismo y no la separación o forma de las nuevas componentes. De hecho se constata umbrales en donde tanto modelos *a priori* y *a posteriori* decrecen su desempeño predictivo al saturar con componentes nuevas.
 - En cuanto al estudio de inclusión de calles de demuestra la flexibilidad de la metodología al poder incorporarse geometrías arbitrarias en el cálculo de riesgo criminal y predicción. Cabe mencionar también la adaptación realizada sobre el HRP y TIP: ya no se evalúa la cantidad de delitos de predicción que cayeron en un determinado segmento de calle, sino el riesgo de predicción proyectado.
- Si bien el algoritmo presentado en esta tesis cumple con los requerimientos del proyecto en conjunto con Carabineros de Chile, es importante realizar comparaciones más rigurosas con los algoritmos que se pueden encontrar en la literatura científica. En ese sentido resulta importante considerar el *framework* desarrollado por [2] y comparar desempeños con el algoritmo desarrollado en esta tesis y las zonas de estudios trabajadas.
 - Las limitaciones de la metodología como se demuestra en Resultados predictivos generales y Resultados de metodología para diferentes comunas, es que el modelo no cuenta con un mecanismo adaptable para modificar el factor inter-procedimiento λ , ni tampoco es simple poder decidir *a priori* cuál versión del algoritmo (estático/dinámico) es el más adecuado para una determinada comuna. Tampoco es claro cómo sería posible agregar información extra como por ejemplo factores demográficos o socio-económicos. Finalmente el algoritmo desde un punto de vista teórico puede ser mejorado incluyendo teoría de procesos Gaussianos (donde la función de intensidad se estimaría diseñando una función de covarianza) y también, desde el punto de vista geométrico, lo ideal sería incluir ecuaciones de calor para estudiar la “transmisión” de riesgo a lo largo de calles.

5.1. Trabajo Futuro

5.1.1. Con Respecto a Métricas de Evaluación Espaciales

Necesariamente se deben explorar nuevas métricas de análisis *online* para tomar decisiones durante el entrenamiento de modelos con respecto a un determinado parámetro, y mejorar la manera de analizar métricas *offline* las cuales se aplican una vez echas pruebas exhaustivas del parámetro a estudiar para generar heurísticas y/o tablas de decisión pre-calculadas. En esta tesis se introduce la métrica *online* de **distancias** para calcular la cantidad de iteraciones dentro de un modelo espacial, y métricas *offline* como las optimizaciones multi-objetivos presentadas para el cálculo del número de componentes óptimas de modelos espaciales. Como trabajo futuro, sería interesante desarrollar una métrica *online* para determinar el número de componentes óptimo del entrenamiento de una GMM.

5.1.2. Propuesta de Método *Online* para Determinar Número de Componentes Óptimos para GMM

Se puede explorar la retroalimentación ya sea ascendente o descendente de las soluciones del entrenamiento de una GMM_N . La alternativa descendente plantea realizar un entrenamiento considerando una cantidad de componentes máxima y a partir de ahí ir decreciendo su número. Debido a que todos los datos pueden ser clasificados a un kernel determinado, pero no necesariamente todos los kernels tienen un dato, una manera de establecer un criterio de término sería ir eliminando en cada iteración los kernels que no tengan un dato asignado, o también mediante un umbral d_{\min} , el cuál establece el número mínimo de datos por kernels para que sobreviva en cada iteración. Por otra parte la alternativa ascendente consistiría en entrenar los modelos a partir de una cantidad mínima de componentes e ir agregando hasta que se cumpla el criterio ya expuesto con un umbral d_{\min} .

5.1.3. Con Respecto a Condiciones Iniciales para GMM

Para evitar costo computacional extra al inicializar aleatoriamente los parámetros para un entrenamiento espacial, se puede realizar en primer lugar un KDE sobre los datos de entrenamiento y buscar las zonas de mayor concentración, las cuales serán tomadas como medias iniciales, mientras que las covarianzas iniciales estarán dadas por una matriz diagonal que refleje al 0.9 de probabilidad una circunferencia de radio de 500 metros. Los pesos iniciales estarán dados por la evaluación de la superficie KDE en los puntos considerados medias iniciales. Cabe mencionar que este procedimiento es más rápido que el procedimiento de escoger como medias iniciales las calles (ya presentado), sin embargo el de calles es más acorde a la fenomenología de tipos de eventos criminales estudiados.

5.1.4. Con Respecto a Mecanismo sobre Valores Propios de Matrices de Covarianza

Actualmente los componentes Gaussianos se subdividen en tres con un cierto grado de separación y con inclinación solidaria al componente original. Sería interesante explorar una subdivisión menos rígida y basada en un algoritmo de *clustering*, como por ejemplo *k-means* y situar ya sea dos o tres componentes con un posicionamiento que dependiente de los datos.

5.1.5. Con Respecto a Algoritmo EM

Todos los entrenamientos espaciales son realizados por medio del algoritmo EM sin ningún tipo de restricción. Sería interesante explorar un algoritmo iterativo para los parámetros de una GMM utilizando EM con restricciones geométricas o sino restricciones no-lineales de desigualdad. Hasta ahora la literatura científica apunta a EM con restricciones lineales con desigualdad, por lo cual una posible extensión para otro tipo de restricciones sería interesante.

5.1.6. *Framework* de Entrenamiento de Parámetros

En la metodología actualmente implementada, todo entrenamiento de parámetros se realiza con un conjunto de datos pasados a partir de la fecha de consulta. En caso de ajustar parámetros se genera una grilla de éstos y se escogen los mejores valores según el mejor AUC para diferentes pasos de predicción. El defecto de esta metodología es que cada consulta de predicción es aislada de las demás, es decir, no basa sus predicciones en otros resultados ya computados. Si el objetivo fuera predecir alguna fecha del 2017 tomando como ejemplo datos de los 3 meses anteriores a partir de la consulta (ver Figura 5.1), se propone haber computado necesariamente las predicciones para las fechas homólogas de los años anteriores guardando los parámetros óptimos. Luego para la fecha del año 2017 a computar, se tendría en promedio cuál debiesen ser aproximadamente los valores óptimos de los parámetros gracias a los análisis computados en los años anteriores (conocimiento *a priori* de los parámetros).

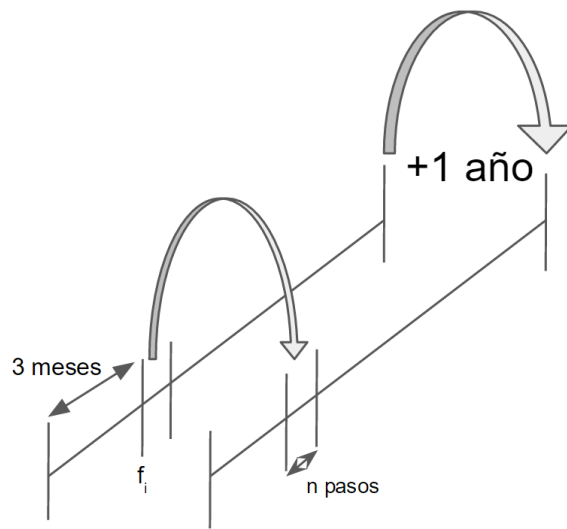


Figura 5.1: *Framework* de entrenamiento sistemático de parámetros.

Bibliografía

- [1] W. Marzocchi, J. Douglas Zechar, and T. H. Jordan, “Bayesian forecast evaluation and ensemble earthquake forecasting,” *Bulletin of the Seismological Society of America*, vol. 102, no. 6, pp. 2574–2584, 2012.
- [2] M. Adepeju, G. Rosser, and T. Cheng, “Novel evaluation metrics for sparse spatio-temporal point process hotspot predictions - a crime case study,” *International Journal of Geographical Information Science*, vol. 0, no. 0, pp. 1–22, 2016.
- [3] D. Duvenaud, “Automatic Model Construction with Gaussian Processes,” Ph.D. dissertation, 2014.
- [4] X. Wang and D. Brown, “The spatio-temporal modeling for criminal incidents,” *Security Informatics*, vol. 1, no. 1, pp. 1–17, 2012.
- [5] J. a. Bilmes, “A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models,” *International Computer Science Institute*, vol. 4, no. 510, p. 126, 1998.
- [6] P. Fuentes, “Diseño y análisis de modelo espacio - temporal para caracterizar y predecir el riesgo criminal usando transformación basada en kernel periódicos gaussiano y funciones wavelet,” *Tesis de Magister (inédidada)*, Universidad de Chile, 2017.
- [7] S. R. Flaxman, “A General Approach to Prediction and Forecasting Crime Rates with Gaussian Processes,” 2014.
- [8] W. L. Perry, B. McInnis, C. C. Price, S. Smith, and J. S. Hollywood, *Predictive Policing*, 2013.
- [9] S. Chainey, L. Tompson, and S. Uhlig, “The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime,” *Security Journal*, vol. 21, pp. 4–28, 2008. [Online]. Available: <http://discovery.ucl.ac.uk/112873/>
- [10] G. O. Mohler, M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita, “Self-Exciting Point Process Modeling of Crime,” *Journal of the American Statistical Association*, vol. 106, no. 493, pp. 100–108, 2011.
- [11] C. Torrence and G. P. Compo, “A practical guide to wavelet analysis,” *Bulletin of the American Meteorological society*, vol. 79, no. 1, pp. 61–78, 1998.

- [12] PredPol, “Predictive policing for crime analysts.” [Online]. Available: <http://www.predpol.com>
- [13] IBM, “Memphis Police Department Reduces Crime Rates with IBM Predictive Analytics Software,” jul 2010. [Online]. Available: <http://www-03.ibm.com/press/us/en/pressrelease/32169.wss>
- [14] UCL, “UCL - London’s Global University,” 2012. [Online]. Available: <http://www.ucl.ac.ukhttp://www.ucl.ac.uk/>
- [15] Fundación Paz Ciudadana, “Balance de la Delincuencia en Chile 2014. Radiografía del sistema de seguridad y justicia,” Tech. Rep., 2015. [Online]. Available: <http://www.pazciudadana.cl/wp-content/uploads/2015/08/balance-de-la-delincuencia-en-chile-fpc-17-agosto-2015.pdf>
- [16] Encuesta CEP, “Estudio nacional de Opinión Pública,” Tech. Rep., 2016. [Online]. Available: https://www.cepchile.cl/cep/site/artic/20160818/asocfile/20160818165239/encuestacep_julio_agosto2016.pdf
- [17] INE, “XII ENCUESTA NACIONAL URBANA DE SEGURIDAD CIUDADANA,” Tech. Rep., 2016.
- [18] Biblioteca del Congreso Nacional de Chile, “Reportes Estadísticos Comunales,” 2015. [Online]. Available: http://reportescomunales.bcn.cl/2015/index.php/Puente_Alto
- [19] O. Figueroa, “PRIMER SEMINARIO NACIONAL DE GEOCIENCIAS Y GEOTECNOLOGIAS , Sistema de Análisis de Información Territorial,” Carabineros de Chile, Tech. Rep., 2015. [Online]. Available: http://cpag.cl/presentaciones-seminario/ppt_CPAG_CARABINEROS.pdf
- [20] W. L. Gorr and K. S. Kurland, *GIS tutorial for crime analysis*. Esri Press Redlands, CA, 2012.
- [21] S. Chainey and J. Ratcliffe, *GIS and crime mapping*. John Wiley & Sons, 2013.
- [22] T. Nakaya and K. Yano, “Visualising crime clusters in a space-time cube: An exploratory data-analysis approach using space-time kernel density estimation and scan statistics,” *Transactions in GIS*, vol. 14, no. 3, pp. 223–239, 2010.
- [23] CONASET. (2013) Análisis espacio temporal de los siniestros de tránsito en el gran santiago. [Online]. Available: <http://www.conaset.cl/wp-content/uploads/2016/08/analisis-espacio-temporal-evolucion-siniestros-GS-2009-2013.pdf>
- [24] K. J. Bowers, S. D. Johnson, and K. Pease, “Prospective hot-spotting: The future of crime mapping?” *British Journal of Criminology*, vol. 44, no. 5, pp. 641–658, 2004.
- [25] P. Flores, “Modelación y predicción de focos de criminalidad basado en modelos probabilísticos,” *Repositorio Académico - Universidad de Chile*, 2014. [Online]. Available: http://repositorio.uchile.cl/bitstream/handle/2250/129832/cf-flores_pa.pdf?

sequence=1

- [26] P. Flores, M. Vergara, P. Fuentes, F. Jaramillo, D. Acuña, A. Perez, and M. Orchard, “Modeling and Prediction of Criminal Activity Based on Spatio-Temporal Probabilistic Risk Functions,” in *Annual Conference of the Prognostics and Health Management Society 2015*, 2015, pp. 1–14. [Online]. Available: <https://www.phmsociety.org/node/1741>
- [27] K. Kempinska, T. Davies, and J. Shawe-Taylor, “Probabilistic map-matching using particle filters,” 2016. [Online]. Available: <http://arxiv.org/abs/1611.09706>
- [28] A. Okabe, T. Satoh, and K. Sugihara, “A kernel density estimation method for networks, its computational method and a GIS-based tool,” *International Journal of Geographical Information Science*, vol. 23, no. 1, pp. 7–32, 2009.
- [29] J. Zhuang, Y. Ogata, and D. Vere-Jones, “Stochastic Declustering of Space-Time Earthquake Occurrences,” *Journal of the American Statistical Association*, vol. 97, no. 458, pp. 369–380, 2002.
- [30] e. Ertekin, C. Rudin, and T. H. McCormick, “Reactive point processes: A new approach to predicting power failures in underground electrical systems,” *Annals of Applied Statistics*, vol. 9, no. 1, pp. 122–144, 2015.
- [31] J. Haworth, J. Shawe-Taylor, T. Cheng, and J. Wang, “Local online kernel ridge regression for forecasting of urban travel times,” *Transportation Research Part C: Emerging Technologies*, vol. 46, pp. 151–178, 2014.
- [32] N. A. Rayner, D. Parker, E. Horton, C. Folland, L. Alexander, D. Rowell, E. Kent, and A. Kaplan, “Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century,” *Journal of Geophysical Research*, vol. 108, no. D14, p. 4407, 2003.
- [33] D. J. Daley and D. Vere-Jones, *An introduction to the theory of point processes.*, 2008, vol. I. [Online]. Available: <http://www.springerlink.com/content/978-0-387-21337-8>
- [34] W. Whitt, “Approximating a Point Process by a Renewal Process, I: Two Basic Methods,” *Operations Research*, vol. 30, no. 1, pp. 125–147, 1982.
- [35] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.
- [36] P. J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis,” *Journal of Computational and Applied Mathematics*, vol. 20, no. C, pp. 53–65, 1987.
- [37] K. Lau and H. Weng, “Climate signal detection using wavelet transform: How to make a time series sing,” *Bulletin of the American Meteorological Society*, vol. 76, no. 12, pp. 2391–2402, 1995.
- [38] A. Dempster, N. Laird, D. Rubin, and Others, “Maximum likelihood from incomplete

data via the EM algorithm,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

Anexos

Tablas Relacionadas con Base de Datos

DELITO	CDELITO	AGRUPADO
FEMICIDIO ART.390 INC.2	720	HOMICIDIO
HOMICIDIO ART. 391 NRO.2	702	HOMICIDIO
HOMICIDIO CALIFICADO ART. 391 NRO. 1	703	HOMICIDIO
HOMICIDIO DE GENDARME EN EL DESEMPEÑO DE SUS FUNCIONES	11101	HOMICIDIO
HOMICIDIO EN RIÑA O PELEA ART. 392	705	HOMICIDIO
INFANTICIDIO ART. 394	707	HOMICIDIO
PARRICIDIO ART. 390 INC.1	701	HOMICIDIO
APROPIACION DE CABLES DE TENDIDO ELECTRICO O DE COMUNICACIONES	849	HURTO
HURTO AGRAVADO (ART. 447 CODIGO PENAL)	826	HURTO
HURTO DE BIENES PERTENECIENTES A REDES DE SUMINISTRO PUBLICO	853	HURTO
HURTO DE HALLAZGO ART. 448	821	HURTO
HURTO FALTA 494 BIS CODIGO PENAL	13028	HURTO
HURTO SIMPLE	801	HURTO
HURTO SIMPLE POR UN VALOR DE 4 A 40 UTM. ART. 432 Y 446 NRO. 2	847	HURTO
HURTO SIMPLE POR UN VALOR DE MEDIA A MENOS DE 4 UTM. A.432 Y 446 NRO.3	848	HURTO
HURTO SIMPLE POR UN VALOR SOBRE 40 UTM. ART. 432 Y 446 NRO. 1	846	HURTO
HURTO SUPERMERCADO O CENTROS COMERCIALES	843	HURTO
CASTRACION Y MUTILACION ART. 395 y 396	718	LESIONES
LESIONES CON ARMA BLANCA	712	LESIONES
LESIONES CON ARMA DE FUEGO	714	LESIONES
LESIONES CON FUEGOS ARTIFICIALES	713	LESIONES
LESIONES GRAVES ART. 397 NRO. 2	709	LESIONES
LESIONES GRAVES GRAVISIMAS ART. 397 NRO. 1	717	LESIONES
LESIONES LEVES ART. 494 NRO. 5	13001	LESIONES
LESIONES MENOS GRAVES ART. 399	710	LESIONES
MUTILACION	719	LESIONES
RINA PUBLICA 496 NRO. 10 CODIGO PENAL	13033	LESIONES
ROBO FRUSTRADO	844	ROBO FRUSTRADO
ROBO DE ACCESORIOS DE VEHICULOS O ESPECIES INTERIOR VEHICULOS	842	ROBO FUERZA
ROBO DE VEHICULO MOTORIZADO ART. 443 INC. 2	831	ROBO FUERZA
ROBO EN BIENES NACIONALES DE USO PUBLICO O SITIOS NO DEST.HAB. ART.443	808	ROBO FUERZA
ROBO EN LUGAR HABITADO O DESTINADO A LA HABITACION ART. 440	809	ROBO FUERZA
ROBO EN LUGAR NO HABITADO ART. 442	810	ROBO FUERZA
ROBO CON CASTRACION, MUTILACION O LESIONES GRAVES GRAVISIMAS A.433 N.1	829	ROBO VIOLENCIA
ROBO CON HOMICIDIO ART. 433 NRO. 1	827	ROBO VIOLENCIA
ROBO CON INTIMIDACION ART. 433, 436 INC. 1 438	802	ROBO VIOLENCIA
ROBO CON RETENCION DE VICTIMAS O CON LESIONES GRAVES ART. 433 NRO. 2	830	ROBO VIOLENCIA
ROBO CON VIOLACION ART. 433 NRO. 1	828	ROBO VIOLENCIA
ROBO CON VIOLENCIA ART. 436 INC. 1 433, 438, 439	803	ROBO VIOLENCIA
ROBO POR SORPRESA ART. 436 INC. 2	804	ROBO VIOLENCIA
VIOLACION	607	VIOLACION
VIOLACION CON HOMICIDIO ART. 372 BIS.	628	VIOLACION
VIOLACION DE MAYOR DE 14 AÑOS ART. 361	637	VIOLACION
VIOLACION DE MENOR DE 14 AÑOS ART. 362	621	VIOLACION

Tabla 5.1: Nombre de delitos y sus agrupados según el tipo DMCS.

Algoritmo EM

Máxima Verosimilitud

Recordando el escenario de Máxima Verosimilitud, se tiene en primer lugar una densidad de probabilidad $p(x|\Theta)$ donde Θ es un set de parámetros de la distribución. También se tiene un conjunto de N vectores muestreados de la distribución anterior, digamos $\mathcal{X} = \{x_1, \dots, x_N\}$. Asumiendo que estas muestras fueron generadas de manera i.i.d. de $p(x|\Theta)$ entonces la verosimilitud se calcula como:

$$\mathcal{L}(\Theta, \mathcal{X}) = p(\mathcal{X}|\Theta) \quad (5.1)$$

$$= \prod_{i=1}^N p(x_i|\Theta) \quad (5.2)$$

La verosimilitud es función de Θ donde \mathcal{X} está fijo. En el problema de Máxima Verosimilitud el objetivo es encontrar el Θ que maximice \mathcal{L} , es decir:

$$\Theta^* = \underset{\Theta}{\operatorname{argmax}} \mathcal{L}(\Theta|\mathcal{X}) \quad (5.3)$$

muchas veces se maximiza $\log \mathcal{L}(\Theta|\mathcal{X})$ para más facilidad en el tratamiento analítico de las expresiones, justificándose el uso de logaritmo puesto que es una función continua y cóncava. Dependiendo de la forma de $p(x|\Theta)$ el problema de optimización puede ser fácil o difícil de resolver. Si la distribución es una distribución Gaussiana unidimensional, tanto μ como σ tienen expresiones cerradas, pero hay otros problemas en donde se requiere el uso de técnicas más elaboradas como el algoritmo de EM.

Pasos de EM

Este algoritmo es de carácter iterativo, el cual encuentra progresivamente los parámetros de una distribución mediante máxima verosimilitud cuando el set de datos de entrenamiento es incompleto. Llamaremos a $\mathcal{Z} = (\mathcal{X}, \mathcal{Y})$ el conjunto de datos completo, \mathcal{X} los datos observados e \mathcal{Y} los datos no observados. Asumiendo la existencia de la distribución conjunta se tiene:

$$p(z|\Theta) = p(x, y|\Theta) = p(y|x, \Theta)p(x|\Theta) \quad (5.4)$$

de esta manera se define la verosimilitud como $\mathcal{L}(\Theta|\mathcal{Z}) = \mathcal{L}(\Theta|\mathcal{X}, \mathcal{Y}) = p(\mathcal{X}, \mathcal{Y}|\Theta)$. El algoritmo EM consta de dos pasos fundamentales, el primero donde se computa una función que corresponde a la esperanza del logaritmo de la verosimilitud condicionado a los datos observados y los parámetros de la iteración anterior (paso E), y luego esa expresión se maximiza para encontrar el set de parámetros siguientes (paso M). Para ello la función a computar es:

$$Q(\Theta, \Theta^{(i-1)}) = \mathbb{E} [\log p(\mathcal{X}, \mathcal{Y}|\Theta) | \mathcal{X}, \Theta^{(i-1)}] \quad (5.5)$$

donde $\Theta^{(i-1)}$ corresponde al cálculo de los parámetros para la iteración $(i - 1)$. Finalmente se resuelve el problema:

$$\Theta^{(i)} = \underset{\Theta}{\operatorname{argmax}} Q(\Theta, \Theta^{(i-1)}) \quad (5.6)$$

En cada iteración se garantiza el incremento del logaritmo de la verosimilitud y por tanto la convergencia de $\Theta^{(i)}$ a un máximo local. Para más detalles sobre el algoritmo revisar [5]. Para su deducción y análisis de convergencia [38].

EM Aplicado al Cálculo de Parámetros de GMM

Para una cantidad finita de componentes Gaussianas, el algoritmo EM tiene una expresión cerrada para el cálculo de parámetros de una GMM que es la siguiente [38]:

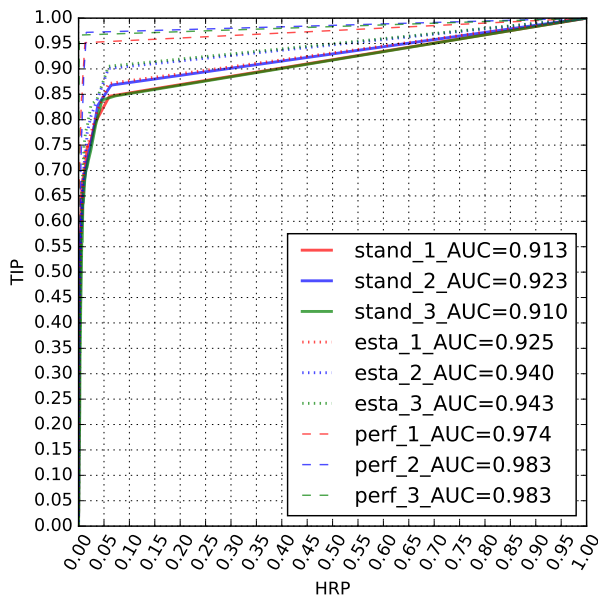
$$w_s^{(j+1)} = \frac{1}{N} \sum_{t=1}^N h_s^{(j)}(t) \quad (5.7)$$

$$\mu_s^{(j+1)} = \frac{\sum_{t=1}^N h_s^{(j)}(t) x^{(t)}}{\sum_{t=1}^N h_s^{(j)}(t)} \quad (5.8)$$

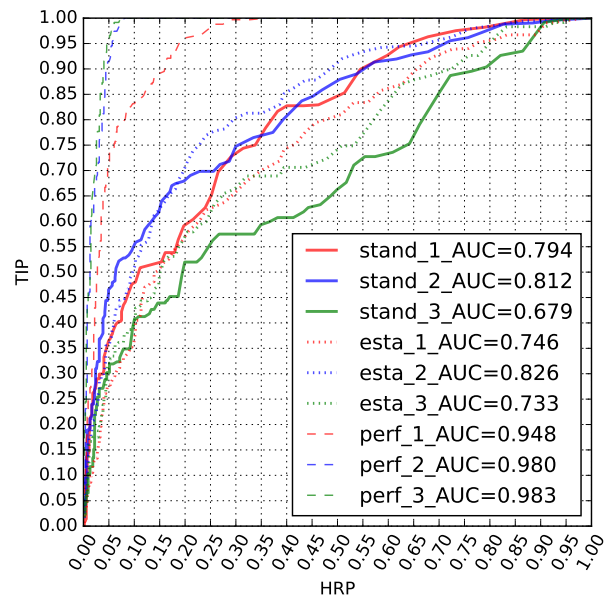
$$\Sigma_s^{(j+1)} = \frac{\sum_{t=1}^N h_s^{(j)}(t) [x^{(t)} - \mu_s^{(j+1)}] [x^{(t)} - \mu_s^{(j+1)}]^T}{\sum_{t=1}^N h_s^{(j)}(t)} \quad (5.9)$$

$$h_s^{(j)}(t) = \frac{w_s^{(j)} p_s(x^{(t)} | \mu_s^{(j)}, \Sigma_s^{(j)})}{\sum_{i=1}^N w_i^{(j)} p_i(x^{(t)} | \mu_i^{(j)}, \Sigma_i^{(j)})} \quad (5.10)$$

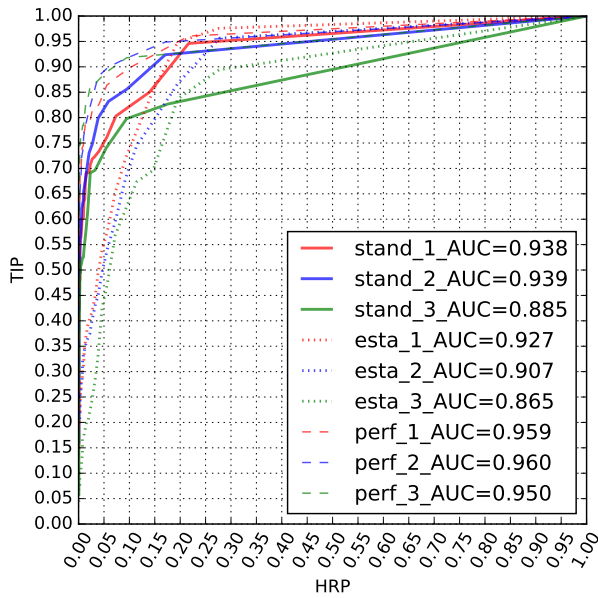
Resultados AUC Promedio para Cada Comuna de Estudio



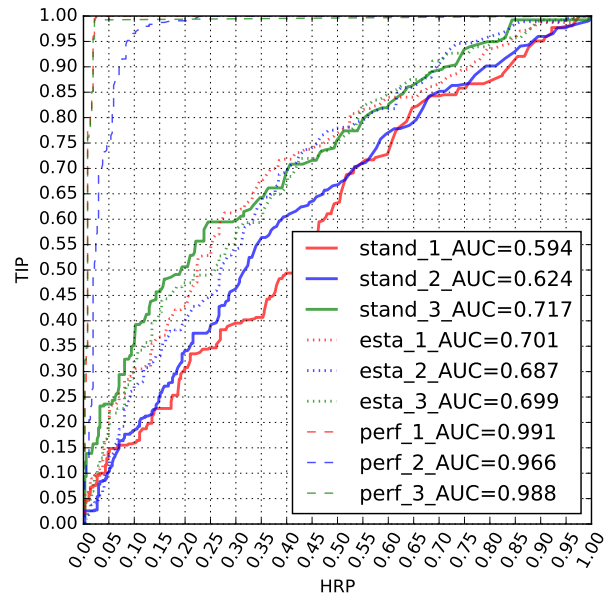
(a) Colina



(b) Conchalí

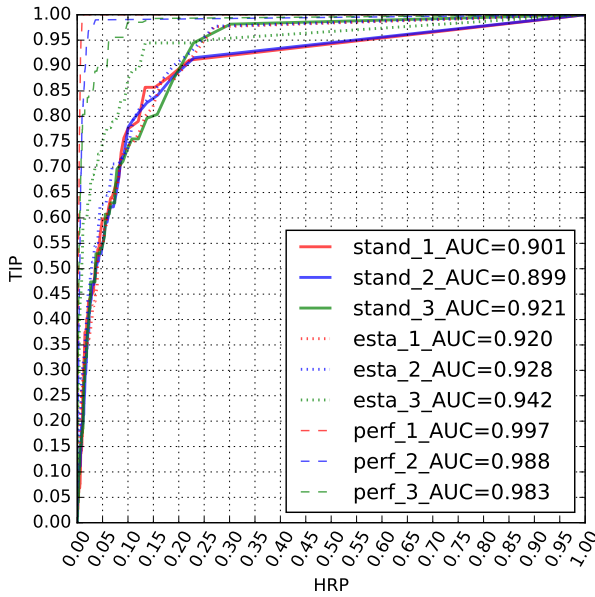


(c) Huechuraba

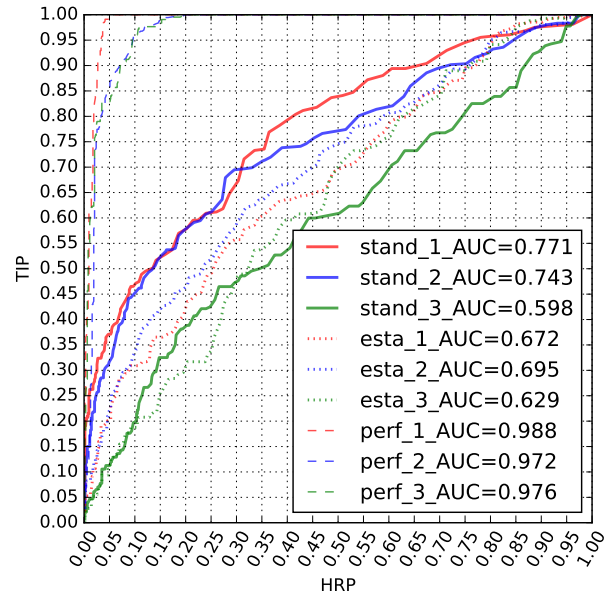


(d) Independencia

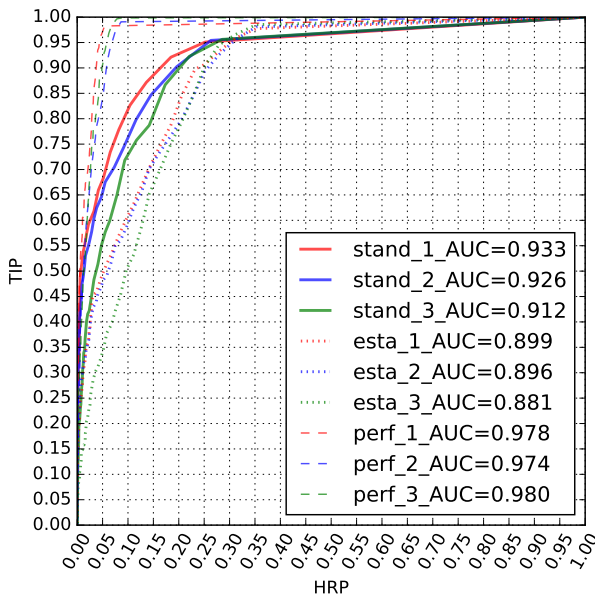
Figura 5.2: AUC's promedio de 1 a 3 pasos de predicción para intervalo de fechas 2016-05-01 2016-08-01 para comunas de estudio.



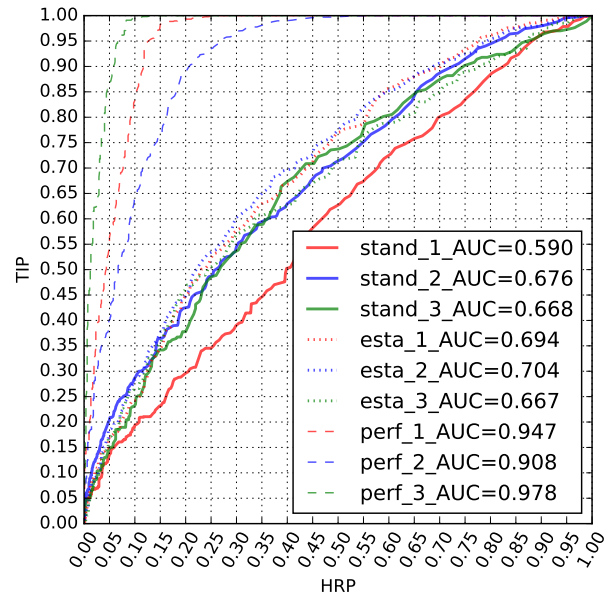
(a) Lampa



(b) Macúl

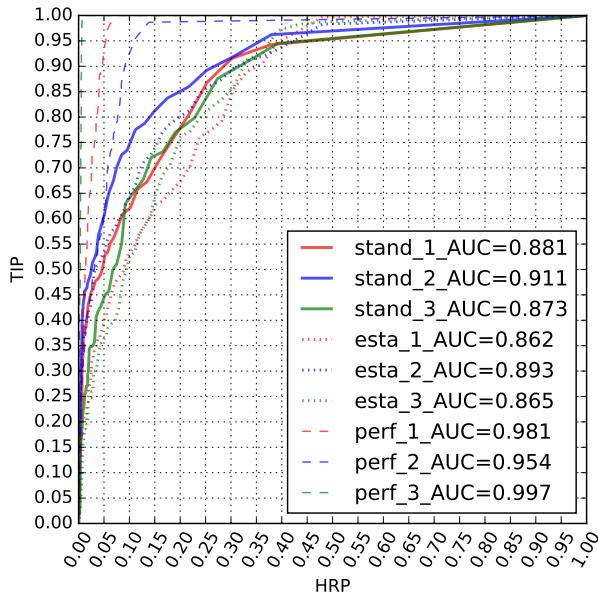


(c) Maipú

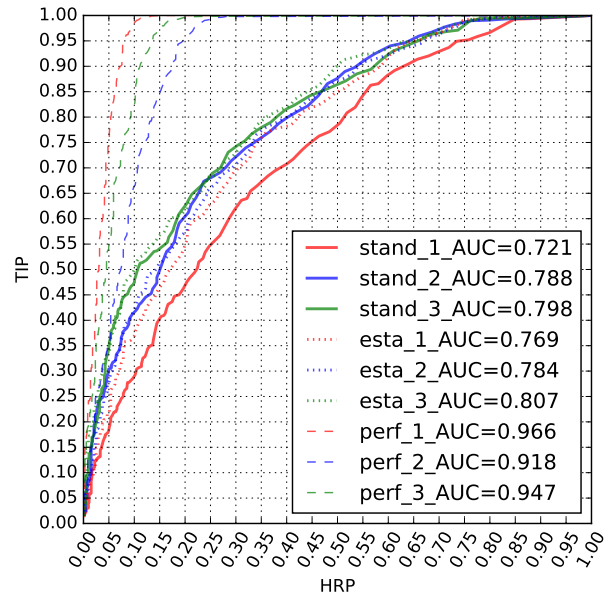


(d) Ñuñoa

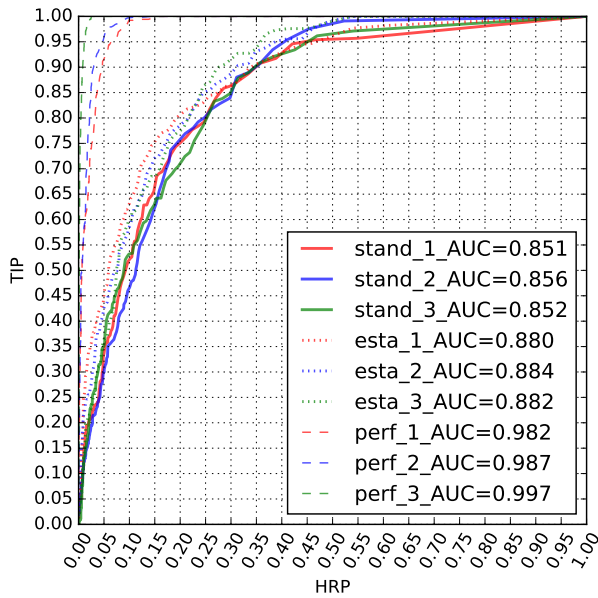
Figura 5.3: AUC's promedio de 1 a 3 pasos de predicción para intervalo de fechas 2016-05-01 2016-08-01 para comunas de estudio.



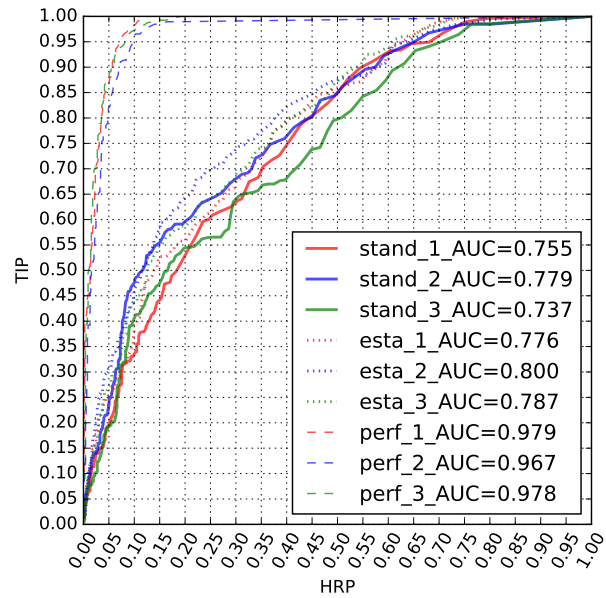
(a) Peñalolén



(b) Providencia

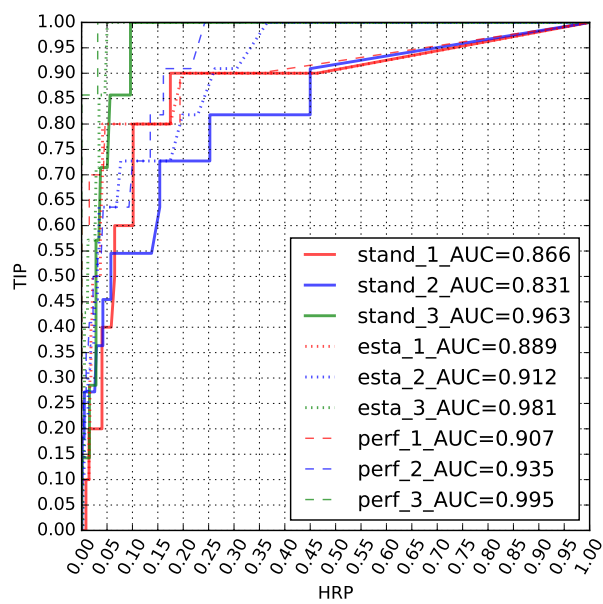


(c) Quilicura



(d) Recoleta

Figura 5.4: AUC's promedio de 1 a 3 pasos de predicción para intervalo de fechas 2016-05-01 2016-08-01 para comunas de estudio considerando las tres versiones del algoritmo.



(a) Tiltit

Figura 5.5: AUC's promedio de 1 a 3 pasos de predicción para intervalo de fechas 2016-05-01 2016-08-01 para comunas de estudio considerando las tres versiones del algoritmo.