



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

**MODELACIÓN Y PREDICCIÓN DE FOCOS DE
CRIMINALIDAD BASADO EN MODELOS PROBABILÍSTICOS**

**MEMORIA PARA OPTAR AL TÍTULO DE INGENIERA CIVIL
ELÉCTRICA**

PAULINA VIVIANA FLORES ARIAS

PROFESOR GUÍA:

MARCOS ORCHARD CONCHA

MIEMBROS DE LA COMISIÓN

RAÚL MANÁSEVICH TOLOSA

JORGE SILVA SÁNCHEZ

SANTIAGO DE CHILE

2014

RESUMEN DE LA MEMORIA PARA OPTAR EL
TÍTULO DE INGENIERA CIVIL ELÉCTRICA.
POR: PAULINA FLORES A.
FECHA: POR DEFINIR
PROF. GUÍA: Dr. MARCOS ORCHARD C.

“MODELACIÓN Y PREDICCIÓN DE FOCOS DE CRIMINALIDAD BASADO EN MODELOS PROBABILÍSTICOS”

Las fuerzas de seguridad requieren modelar patrones de riesgo asociados a la actividad delictual, para estudiar relaciones de causalidad y estimar dónde, y cuándo, un nuevo evento criminal puede ocurrir. Esta información se vuelve muy relevante a la hora de asignar recursos para mejorar la seguridad y evitar la ocurrencia de los delitos, y muchos modelos han sido creados con este propósito mediante una caracterización espacial de las funciones de riesgo. Lamentablemente, dado que la evolución temporal del sistema delictual no se encuentra plenamente incorporada en el modelo del riesgo, en la mayor parte de los casos dichos modelos se vuelven obsoletos. En este Trabajo de Título se presenta un método empírico para la generación automática de funciones probabilísticas espaciales de riesgo, junto con un mecanismo para la caracterización de la evolución temporal de los focos de criminalidad. Este método utiliza información georeferenciada de incidentes criminales y servicios para aproximar la distribución espacial del riesgo a través de la suma de kernels Gaussianos. Por su parte, la caracterización de la evolución temporal del modelo se realiza mediante un algoritmo basado en los métodos secuenciales de Monte Carlo (también conocidos como filtros de partículas), incorporando recursivamente nuevas observaciones y aproximando una distribución mediante la posición de partículas en el plano. La utilización de datos reales para caracterizar el riesgo en cada coordenada del plano y las actualizaciones recursivas al incorporar más información, permiten estimar una distribución a posteriori que implícitamente incorpora la variable temporal. El modelo de riesgo obtenido puede utilizarse para definir un modelo de la predicción, actualizándolo con nuevos datos en el corto y largo plazo y modelando cambios en la distribución a medida que el entorno también cambia. Los resultados obtenidos con este método permiten afirmar que la mayor parte de los eventos futuros ocurren dentro de las áreas de riesgo modeladas, teniendo entonces el potencial de asistir en la asignación de recursos, mejorando los planes de intervención en las zonas de riesgo estimadas (labor táctica) y predichas (labor estratégica).

AGRADECIMIENTOS

Finalizando esta importante etapa de mi vida, quisiera tomarme el tiempo de agradecer a aquellos que me acompañaron y apoyaron durante este proceso.

En primer lugar agradezco a mis padres, por siempre incentivar en mí el pensamiento crítico y el espíritu de superación, y agradezco cada uno de sus sacrificios por enseñarme y guiarme en este mundo bastante loco.

Agradezco a mi hermana por siempre estar presente, porque su fe en mí constantemente me hizo querer ser mejor, y no conformarme ni quedarme quieta esperando a que las cosas pasen. Tenerte como hermana mayor me hizo siempre seguir tus pasos, llevándome hasta donde estoy ahora.

Quisiera agradecer a mi pololo, por su paciencia y sus consejos, por cada una de las veces que fue la única persona en soportar mi locura, y por nunca perder las esperanzas de esta vida que queremos construir juntos. Sé que esta etapa hubiese sido mucho más difícil sin él, y su presencia en mi vida me hace perderle el miedo a todo lo que se viene en el futuro.

Agradezco también a todos mis amigos, por hacer más soportable el caos de la universidad, por siempre tener un saludo amable y por aceptarme con cada uno de mis defectos. Por compartir alegrías y decepciones, buenos y malos momentos, por todas esas veces que me ayudaron mucho más de lo que supieron, por hacerme saber que no estaba sola en esta etapa, por cada una de sus palabras y por la confianza que depositaron en mí y que me dejaron depositar en ellos.

Quisiera agradecer también a aquellos pequeños y silenciosos integrantes de la familia, pues cada vez que los miro me hacen darme cuenta que las cosas no son tan complicadas como parecen.

Agradezco a todos aquellos que ayudaron en la realización de este trabajo, en especial a los profesores miembros de esta comisión; al Profesor Guía por su constante y fundamental apoyo, por sus conocimientos y por su disposición; y a Carabineros de Chile por facilitar la información en la que se basa este trabajo.

TABLA DE CONTENIDO

CAPÍTULO 1. INTRODUCCIÓN.....	1
1.1 MOTIVACIÓN.....	1
1.2 ALCANCE.....	3
1.3 OBJETIVOS	4
1.4 INDICACIÓN SOBRE CONFIDENCIALIDAD	4
1.5 ESTRUCTURA GENERAL.....	5
CAPÍTULO 2. REVISIÓN BIBLIOGRÁFICA Y ESTADO DEL ARTE	6
2.1 ANÁLISIS Y MAPEO DE CRÍMENES.....	6
2.2 MODELOS DE HOT SPOTS.....	9
2.3 ACTIVIDAD POLICIAL PREDICTIVA	11
2.4 MODELOS PROBABILÍSTICOS PARA INCIDENTES CRIMINALES.....	12
2.5 MODELOS POR MEZCLA DE DISTRIBUCIONES (MIXTURE MODELS)	15
2.6 INFERENCIA BAYESIANA E INTEGRACIÓN DE MONTE CARLO.....	17
2.7 MUESTREO DE IMPORTANCIA (IMPORTANCE SAMPLING).....	19
2.8 REMUESTREO (RESAMPLING).....	20
2.9 FILTRO DE MÍNIMOS CUADRADOS RECURSIVO.....	22
2.10 MÉTODO DE EVALUACIÓN DE MODELOS.....	25
CAPÍTULO 3. MÉTODO DE GENERACIÓN AUTOMÁTICA DE DISTRIBUCIONES ESPACIALES DE RIESGO Y CARACTERIZACIÓN DE SU EVOLUCIÓN TEMPORAL....	28
3.1 INFORMACIÓN REQUERIDA	29

3.1.1	DEFINICIÓN DEL ÁREA DE INTERÉS.....	29
3.1.2	DATOS GEOREFERENCIADOS DE SERVICIOS ASOCIADOS AL ÁREA DE INTERÉS.....	30
3.1.3	DATOS GEOREFERENCIADOS DE EVENTOS DELICTUALES ASOCIADOS AL ÁREA DE INTERÉS	31
3.2	GENERACIÓN DE MODELOS ESPACIALES DE RIESGO	31
3.2.1	DESCRIPCIÓN GENERAL DEL ESQUEMA	32
3.2.2	DESCRIPCIÓN DETALLADA DEL ESQUEMA.....	34
3.3	CARACTERIZACIÓN DE LA EVOLUCIÓN TEMPORAL DE LA DISTRIBUCIÓN ESPACIAL DE RIESGO	37
3.3.1	DESCRIPCIÓN GENERAL DEL ESQUEMA	38
3.3.2	DESCRIPCIÓN DETALLADA DEL ESQUEMA.....	41
CAPÍTULO 4. CASO DE ESTUDIO Y DISCUSIÓN DE RESULTADOS OBTENIDOS		57
4.1	DESCRIPCIÓN DE LA COMUNA ESTUDIADA Y DE LAS BASES DE DATOS UTILIZADAS	57
4.2	RESULTADOS REFERENTES A LA METODOLOGÍA DE GENERACIÓN AUTOMÁTICA DE MODELOS ESPACIALES DE RIESGO	60
4.3	RESULTADOS REFERENTES A LA METODOLOGÍA DE CARACTERIZACIÓN DE LA EVOLUCIÓN TEMPORAL DE MODELOS ESPACIALES DE RIESGO	66
CAPÍTULO 5. CONCLUSIONES Y TRABAJO FUTURO		83
BIBLIOGRAFÍA		86

ÍNDICE DE FIGURAS

Figura 1: Relación entre análisis y mapeo de crímenes. El mapeo corresponde a una disciplina del análisis, pudiendo asistir en los tres niveles de éste: nivel táctico, estratégico y administrativo. ...	8
Figura 2: Diferencias entre las definiciones de Hot Spot. En A se aprecia teoría de lugares; en B la teoría de calles; en C la teoría de vecindarios.....	10
Figura 3: Representación de un GMM de M componentes. A partir de un vector de estados, se definen M distribuciones gaussianas, cada una con su vector de medias y matriz de covarianzas, y mediante una suma ponderada es posible obtener el modelo definitivo.	17
Figura 4: Operación del algoritmo de resampling utilizado.	21
Figura 5: Esquema del algoritmo RLS. El error entre la señal deseada y la señal estimada por un filtro (variable) es utilizado para actualizar los parámetros del mismo, de forma recursiva cada vez que una nueva medición se encuentra disponible.	22
Figura 6: Esquema general del método propuesto para la generación de distribuciones espaciales de riesgo y caracterización de su evolución temporal.	29
Figura 7: Representación de la relación entre el área de interés y la cantidad de servicios asociados. Se aprecia que el uso de pocos servicios no representará fielmente todas las áreas de riesgo, y el uso de muchos servicios solo produce un aumento en el gasto computacional.....	30
Figura 8: Diagrama de flujo general de la metodología de generación automática de modelos de riesgo.	32
Figura 9: Representación de la localización de eventos (A), eventos y servicios (B), y eventos, servicios y el rango para cada uno de ellos (C). Solo corresponde a una representación y muy probablemente no todos los eventos delictuales se ubiquen cercanos a los servicios.....	33
Figura 10: Diagrama de flujo detallado de la metodología de generación automática de modelos de riesgo.....	34
Figura 11: Ejemplo de las curvas de nivel de un modelo hipotético obtenido al final del método propuesto. Cada una de las distribuciones gaussianas se encuentra centrada en la posición del servicio que generó la zona de riesgo, orientándose automáticamente según la distribución original de los incidentes, y no de acuerdo al rango considerado (es decir, no todas las distribuciones serán iguales).....	37
Figura 12: Diagrama de flujo general de la metodología de caracterización de la evolución temporal.....	39

Figura 13: Ejemplo del umbral de operación del modelo dinámico. Del total de partículas en un área de interés dada (A), tras la definición de un umbral de operación solo algunas de ellas serán afectadas por la observación (B).	40
Figura 14: Ejemplo del movimiento de las partículas por efecto de la observación y del modelo dinámico. Se aprecia que aquellas partículas lejanas (fuera del umbral de operación) se desplazan solo por efecto del ruido (en verde y azul), mientras que aquellas ubicadas dentro del umbral de operación experimentarán un desplazamiento mayor generado por la atracción que genera la observación (gris y negro).	41
Figura 15: Diagrama de flujo detallado de la metodología offline de caracterización de la evolución temporal y predicciones del riesgo espacial.	42
Figura 16: Ejemplo de muestreo de importancia y resampling. En A se aprecian las partículas obtenidas mediante el muestreo de importancia, que distribuye uniforme en el área de interés. En B se aprecian las partículas obtenidas después del algoritmo de resampling, presentando una mayor densidad en las zonas de mayor probabilidad de acuerdo a la distribución a priori.	43
Figura 17: Diagrama de flujo detallado de la metodología online de caracterización de la evolución temporal y predicciones.	48
Figura 18: Diagramas de flujo de los Métodos 1 y 2, implementados para la predicción.	49
Figura 19: Ejemplo de la caracterización del movimiento de una partícula afectada por la observación mayoritariamente. En A se presenta el desplazamiento real que experimenta una partícula dadas sucesivas observaciones; estos desplazamientos secuenciales se expresan en coordenadas polares, presentándose el módulo o magnitud del movimiento y el ángulo u orientación, en B y C respectivamente.	50
Figura 20: Esquema de las transformaciones y geometría del Método 1. En A se muestra el desplazamiento total que ha experimentado una partícula dada, en coordenadas cartesianas, y luego expresado en coordenadas polares en B. Mediante dos filtros RLS es posible obtener una predicción para la magnitud y la orientación del próximo movimiento presentado en C, y esta nueva posición predicha se transforma al plano cartesiano nuevamente en D.....	52
Figura 21: Representación de criterio de atenuación para el Método 1. En rojo se presentan los valores reales que se estiman para la posición futura de la partícula, tanto en el plano polar (A) como cartesiano (B). En verde se presentan los valores atenuados para la posición futura de la partícula, tanto en el plano polar (A) como cartesiano (B).	53

Figura 22: Ejemplo de la caracterización del movimiento de una partícula afectada por el ruido mayoritariamente. La representación del movimiento futuro en el plano cartesiano (A) puede fácilmente transformarse al plano polar (B), donde conservando el valor de la orientación se atenúa la magnitud del movimiento (C), y la posición final atenuada se transforma nuevamente al plano cartesiano (D).....	55
Figura 23: Comuna de La Florida, presentando su ubicación dentro de la ciudad de Santiago de Chile.	58
Figura 24: Límites comunales de La Florida y sus avenidas principales.	60
Figura 25: Localización de incidentes delictuales y de servicios en el área de interés. A simple vista no es posible identificar focos de criminalidad, así como tampoco se aprecia una clara correlación entre ambos.....	61
Figura 26: Modelos generados para servicios tipo 'Botillería', considerando 3, 5 y 7 cuadras respectivamente para su creación.	61
Figura 27: Modelos generados para servicios del tipo 'Gran Tienda Comercial', considerando 3, 5 y 7 cuadras respectivamente para su creación.....	62
Figura 28: Modelos generados para servicios del tipo 'Bar y Pub', considerando 3, 5 y 7 cuadras respectivamente para su creación.	62
Figura 29: Gráficos comparativos de los tiempos de procesamiento para la generación del modelo, considerando un rango de 3, 5 y 7 cuadras respectivamente. No se aprecia una diferencia sustancial en ellos.	63
Figura 30: Distribución espacial del riesgo, considerando los valores que optimizan criterio de AUC y distribución de recursos: rango 6 cuadras y grilla de 8 cuadras.....	65
Figura 31: Esquema representativo del modelo de riesgo generado (superficie en escala de colores de blanco a amarillo y rojo) y la distribución espacial de los eventos del conjunto de prueba (cuadrados azules). Se aprecia que la mayoría de estos eventos se presentan en las zonas catalogadas como "de alto riesgo" o cerca las mismas.....	66
Figura 32: Comparación entre posiciones de 100 partículas en base a una distribución uniforme y posterior al remuestreo.	67
Figura 33: Comparación entre posiciones de 250 partículas en base a una distribución uniforme y posterior al remuestreo	68

Figura 34: Comparación entre posiciones de 500 partículas en base a una distribución uniforme y posterior al remuestreo	68
Figura 35: Posición de partículas, donde se puede apreciar que no existe representatividad en el foco de zona norponiente.....	69
Figura 36: Posición de partículas, donde se puede apreciar que existe escasa representatividad en la zona central.....	69
Figura 37: Comparación entre posiciones de 96 partículas en base a grilla de 10 cuadras y posterior al remuestreo.	70
Figura 38: Comparación entre posiciones de 247 partículas en base a grilla de 6 cuadras y posterior al remuestreo.	70
Figura 39: Comparación entre posiciones de 551 partículas en base a grilla de 4 cuadras y posterior al remuestreo.	70
Figura 40: Comparación entre posiciones de 368 partículas en base a grilla de 5 cuadras y posterior al remuestreo	71
Figura 41: Comparación entre la distribución a priori del riesgo y su aproximación mediante muestreo de importancia (IS) y modelo por mezcla de gaussianas (GMM).	72
Figura 42: Comparación entre el ajuste producido por diferentes varianzas. En la imagen de la izquierda se aprecian las posiciones iniciales de las partículas; en la imagen central se aprecia el efecto de una matriz de covarianza de valores muy altos, produciendo un acercamiento de todas las partículas independiente de su lejanía a la observación; en la imagen de la derecha se aprecian los efectos de una matriz de covarianza con valores muy pequeños, sin notarse un claro desplazamiento incluso de las partículas cercanas.	73
Figura 43: Ajuste en la posición de las partículas producido por una varianza apropiada. En la imagen de la izquierda se muestra la posición inicial de las partículas, mientras que en la imagen de la derecha se aprecian los efectos del uso de una matriz de covarianza apropiada: las partículas lejanas se mantienen lejos, y las cercanas se acercan aún más a la observación.....	73
Figura 44: Ejemplo del movimiento de las partículas afectadas por la presencia de una observación. Comparando la imagen de la izquierda con las dos siguientes, se puede apreciar el movimiento y acercamiento de las partículas en las zonas de interés marcadas con un círculo rojo.	74

Figura 45: AUC para los modelos obtenidos mediante la metodología de evolución temporal con el modelo dinámico propuesto. Se aprecia que incluso en los momentos en que se alcanzan los valores mínimos, estos siguen siendo superiores a 0.83, demostrando así que el modelo sigue representando de buena forma los incidentes futuros.....	75
Figura 46: Esquema comparativo entre la distribución espacial de riesgo (amarillo y rojo) posterior al algoritmo de evolución temporal propuesto y eventos futuros del conjunto de prueba (azul). Se aprecia que muchos eventos continúan ocurriendo en las zonas de alto riesgo, pero a diferencia del modelo presentado en la Figura 31 se comienzan a apreciar nuevos focos en las zonas no caracterizadas previamente.....	76
Figura 47: Ejemplo de trayectoria en coordenadas polares de una partícula que es afectada por la observación mayoritariamente. Se aprecia un claro aumento en la distancia entre la posición inicial y final de la partícula en cada iteración, indicando que se está alejando de su lugar de inicio para “explorar” otra zona del mapa en la que ocurren las observaciones recientes, mostrándose además una clara convergencia de la orientación del movimiento.	77
Figura 48: Ejemplo de trayectoria en coordenadas polares de una partícula que es afectada por el ruido de proceso mayoritariamente. La distancia entre la posición inicial y final en cada iteración se mantiene relativamente constante (notar orden de magnitud), creciendo y decreciendo con cada nueva observación. Esto indica que se mantiene cerca de su posición inicial, sin apreciarse tampoco tendencias en la orientación del movimiento.....	78
Figura 49: Ejemplo de la predicción a un paso mediante RLS para una partícula cualquiera. Considerando los movimientos previos (no mostrados en la imagen) y la última observación disponible se obtiene una predicción de la posición futura.....	79
Figura 50: Ejemplo de la evolución temporal del modelo y predicción a un paso. Es posible apreciar en el acercamiento de cada imagen el movimiento de las partículas ubicadas dentro del umbral de acción de la última observación adquirida, apreciando una acumulación de estas y con ello un aumento del riesgo en esa zona. No son notorios los efectos de la atenuación del movimiento ni del aumento de la incertidumbre.....	79
Figura 51: Ejemplo de la predicción a 6 y 7 pasos. Aquí ya es posible apreciar los efectos de la atenuación del movimiento y aumento de la incertidumbre, que se traducen en cambios poco notorios en la distribución del riesgo en la zona de interés.....	80

Figura 52: Cambios secuenciales del indicador AUC para las predicciones generadas mediante la metodología propuesta. A medida que se realiza predicción a más pasos se aprecia cada vez cambios más sutiles en el indicador entre un tiempo y otro, debido a los efectos de la atenuación y aumento de la incertidumbre. En ningún caso se obtienen valores inferiores a 0,8 indicado que en cada serie los modelos generados representan fielmente los incidentes futuros y por ende el riesgo en el área de interés..... 81

Figura 53: Esquema comparativo entre las predicciones de la distribución de riesgo y eventos futuros del conjunto de prueba. Nuevamente es posible apreciar la relación existente entre las zonas caracterizadas como “de alto riesgo” y la ocurrencia de eventos delictuales en ellas. 82

ÍNDICE DE TABLAS

Tabla 1: Proporción entre tasa de aprehensiones en flagrancia y tasa de denuncias formales para los delitos de mayor connotación social en Chile, ambas expresadas en casos cada 100 mil habitantes	2
Tabla 2: AUC para los modelos generados, variando el rango de modelación de 3 a 7 cuadras, y el tamaño de la grilla de 1 a 10 cuadras.....	64
Tabla 3: AUC para las predicciones del modelo probabilístico.	81

CAPÍTULO 1. INTRODUCCIÓN

1.1 MOTIVACIÓN

De acuerdo al Estudio Nacional de Opinión Pública N° 70 del Centro de Estudios Públicos de Chile, realizada durante los meses de septiembre y octubre del año 2013 [1], la delincuencia, los asaltos y los robos son uno de los principales problemas a los que el gobierno debiese dedicar mayor esfuerzo en solucionar. Sin embargo, es justamente en esto temas donde resulta peor evaluado.

La criminalidad afecta a la comunidad de dos grandes formas: la victimización en sí y la percepción de inseguridad. Ambas se encuentran muy relacionadas, principalmente porque la delincuencia no afecta solo a la víctima efectiva del delito, sino que genera inseguridad en otras personas que no han sido victimizadas. Así, de acuerdo al Balance de Delincuencia 2012, realizado por la Fundación Paz Ciudadana[2], entre los años 2011 y 2012 el porcentaje de hogares que indican tener un grado alto de temor disminuye en hogares que han sido víctima de delitos, mientras que se ha producido un aumento del temor en los hogares que no han sido victimizados.

Además, de acuerdo a la Encuesta Nacional Urbana de Seguridad Ciudadana (ENUSC) del año 2011[3], se muestran importantes incrementos en todos los delitos, siendo las lesiones y el robo con fuerza en la vivienda los que muestran las mayores alzas (41,9% y 33,8% respectivamente). También, los hogares que declararon haber sido revictimizados se incrementaron en 15,5%, confirmando que la cantidad de delitos cometidos en el país también aumentó. Por su parte, las denuncias registraron un ligero crecimiento de 3,7%; la reacción frente al delito tiene un aumento del 10,8%; y se acompaña del fuerte crecimiento de la percepción de que la delincuencia aumenta tanto a nivel país como del entorno cercano (38,5% a nivel país, 31,5% a nivel comuna y 24,7% a nivel barrio). También de acuerdo a esta encuesta, aumenta significativamente la percepción de exposición frente al delito.

La existencia de numerosos delitos en la vía pública y los recursos policíacos limitados hacen necesario un alto desempeño por parte de carabineros para impedir el crimen. Sin embargo, las estadísticas muestran que el desempeño de carabineros puede mejorar.

La Tabla 1 muestra la proporción entre denuncias formales (que la ciudadanía realiza en alguna unidad policial posterior a la ocurrencia del delito) y aprehensiones en flagrancia (mientras ocurre el ilícito) de los delitos de mayor connotación social (DMCS) en Chile; ambas corresponden a la tasa anual cada 100 mil habitantes [4]. La tendencia de la gente a denunciar cada vez más los delitos de los que son víctima genera información de mejor calidad para la policía en relación a los sucesos delictivos. Sin embargo, el uso de esta información no ha generado una mejora de la misma magnitud en la detención de los culpables. Esto considerando patrones de delitos relativamente persistentes en el espacio-tiempo que las estrategias de la policía no han podido erradicar.

Tabla 1: Proporción entre tasa de aprehensiones en flagrancia y tasa de denuncias formales para los delitos de mayor connotación social en Chile, ambas expresadas en casos cada 100 mil habitantes

	2005	2006	2007	2008	2009	2010	2011	2012	2013
<i>Denuncias</i>	2.502,0	2.489,5	2.667,7	2.714,7	2.889,7	2.780,3	3.010,1	2.720,4	2.730,1
<i>Aprehensiones</i>	728,6	735,6	828,1	860,0	946,7	905,2	979,0	896,8	842,1
<i>$\frac{\text{Aprehensiones}}{\text{Denuncias}}$</i>	0,291	0,295	0,310	0,317	0,328	0,326	0,325	0,330	0,308

Además, según la encuesta "Delincuencia: La Percepción de la Gente", realizada por el Instituto Libertad y Desarrollo [5], en Chile se cometen más de 10.000 delitos por día, haciéndose necesario contar con acciones policiales que sean más inteligentes, eficaces y proactivas, a diferencia del sistema actual basado en una reacción a los actos delictuales.

Si bien existen muchos métodos para ayudar en la respuesta de la policía frente al crimen, la predicción de dónde y cuándo cierto delito tiene mayor probabilidad de ocurrir ha ganado considerable interés en el último tiempo. Ahora, se está bastante lejos del ideal de predicción relacionado a la ocurrencia de un delito específico en lugar y tiempo, pero dado que el ser humano es un ser de costumbres y la ocurrencia de delitos generalmente se produce dentro de los

llamados focos o “Hot Spots”, la modelación y predicción de estos focos representa un acercamiento al caso ideal, motivando así la realización de este Trabajo de Título.

1.2 ALCANCE

Este Trabajo de Título contempla el diseño de una metodología que permita establecer de forma cuantitativa las prioridades de asignación de recursos policíacos en una zona específica, entregando un modelo espacial de riesgo de dicho terreno y modificaciones secuenciales de este en el tiempo a medida que el entorno cambia. Se desarrolla también una metodología empírica para obtener una predicción de los focos de riesgo, asistiendo así en el corto y largo plazo.

El esquema utiliza información georeferenciada de eventos delictuales y servicios asociados a un área en particular tanto para la creación de modelos espaciales de riesgo como para caracterizar su comportamiento en el tiempo. Se propone entonces una metodología basada en el caso de estudio de la comuna de La Florida para desarrollar un esquema de modelación y predicción potencialmente escalable, tomando como base los delitos catalogados como robo con fuerza.

En este Trabajo de Título no se considera una implementación real y puesta en marcha del método propuesto, pues ésta representa una primera aproximación a la actividad predictiva de las policías y se requiere de un mayor análisis por parte de expertos para pensar en su implementación real. Tampoco se pretende generalizar este modelo, puesto que muchos parámetros utilizados están basados en el caso empírico con el cual se trabaja y en caso de estudiar la escalabilidad de la propuesta se requerirá el mismo nivel de análisis para la obtención de estos parámetros que el utilizado durante la realización del Trabajo de Título. Además, para el desarrollo de este no se cuestionan los datos proporcionados, asumiéndose entonces que todos estos son correctos y que no se han presentado problemas al momento de ingresarlos a la base de datos utilizada (por ejemplo, errores de escritura o valores imprecisos). Tampoco se aborda la distribución de recursos en sí, dejándose esto a manos de expertos en el tema que puedan utilizar la metodología propuesta.

1.3 OBJETIVOS

El objetivo principal de este Trabajo de Título es la caracterización y predicción de focos de criminalidad basada en modelos probabilísticos, con el fin de establecer prioridades en la asignación de recursos policiales en un área determinada y asistir a nivel táctico y estratégico a corto y largo plazo.

Para lograr lo anterior se ha definido un conjunto de objetivos específicos que permiten realizar tal tarea, y representan los hitos más relevantes en los cuales se desarrolla el trabajo:

- Diseño de una metodología empírica de modelación de la distribución espacial de riesgo dentro de un área determinada, a partir de información georeferenciada de eventos delictuales y servicios.
- Diseño de una metodología empírica de caracterización de la evolución temporal de la distribución espacial de riesgo mediante inclusión secuencial de nueva información al modelo inicial.
- Diseño de una metodología empírica de predicción de los focos de criminalidad mediante análisis secuencial de datos y filtro RLS.
- Aplicación y validación de las metodologías propuestas a un caso de estudio.

1.4 INDICACIÓN SOBRE CONFIDENCIALIDAD

Este Trabajo de Título está inserto en un proyecto de Carabineros de Chile en colaboración con la institución CEAMOS. CEAMOS (*Center for Analysis and Modeling of Security*) es un centro de investigación aplicada para el análisis y modelamiento de la seguridad, cuyo objetivo es generar un centro de excelencia de nivel internacional, que desarrolle investigación de punta en seguridad y que nutra las decisiones de política en estas materias [6]. Debido a la presencia de acuerdos de confidencialidad entre las distintas partes, existe información relevante referida a los algoritmos desarrollados y las bases de datos utilizadas que no será entregada detalladamente en este documento. Sin embargo, se deja constancia en este documento que el Profesor Guía y Co-guía están en conocimiento de toda la información no presentada en este documento.

1.5 ESTRUCTURA GENERAL

Este Trabajo de Título está conformado por cinco capítulos.

En el Capítulo 1 se realiza una pequeña introducción al tema a tratar.

En el Capítulo 2 se presenta la revisión bibliográfica y el estado del arte sobre las metodologías utilizadas actualmente. Se describen también distintos conceptos necesarios para comprender el contexto de este trabajo, y se presenta un marco teórico sobre las diferentes técnicas utilizadas para la realización del mismo: modelos por mezcla de gaussianas, muestreo de importancia, filtro RLS, entre otros.

El Capítulo 3 describe las propuestas para la modelación, caracterización temporal y predicción que han sido implementadas. Primero se describe la información que se requiere para utilizar las técnicas propuestas, luego para cada una de las metodologías se entrega una descripción general y una descripción detallada. Esto con la finalidad de introducir al lector a la comprensión del espíritu de la propuesta para luego adentrarse en los detalles de la misma.

El Capítulo 4 presenta la aplicación de las metodologías propuestas a un caso de estudio, el cual es utilizado para probar y validar los esquemas de modelación y predicción descritos en el capítulo anterior. Además se entregan los resultados obtenidos en dicho estudio, los cuales son presentados de forma separada para cada una de las etapas que constituyen la propuesta. Análisis de tales resultados son también presentados en este capítulo.

Para finalizar, el Capítulo 5 entrega las conclusiones finales del Trabajo de Título, además de proponer recomendaciones para el trabajo futuro.

CAPÍTULO 2. REVISIÓN BIBLIOGRÁFICA Y ESTADO DEL ARTE

A lo largo y ancho del globo, las fuerzas de seguridad y orden se dedican a monitorear los incidentes criminales que se producen en un área y rango de tiempo determinado, y gracias a nuevas tecnologías de localización se ha hecho posible además la inclusión de información georeferenciada. Los analistas involucrados buscan entonces patrones espacio-temporales en los incidentes con el fin de predecir cuándo y dónde se producirá un nuevo delito. Esta predicción busca optimizar las actividades que realizan las policías e identificar posibles objetivos para intervención policial, cuyo objetivo fundamental es evitar la ocurrencia de un crimen. Así, se requiere de un modelo que sea capaz de incorporar tanto la información geográfica como temporal para caracterizar la actividad delictual y realizar esta predicción. Estos métodos permiten a la policía trabajar de forma más proactiva con los recursos limitados que poseen, buscando siempre evitar la ocurrencia de estos eventos.

En esta sección se explican los contenidos necesarios para comprender la investigación y el trabajo realizado. Se comienza con definiciones y conceptos ampliamente utilizados en esta materia, para luego seguir con conocimientos técnicos específicos utilizados para la implementación del esquema de caracterización.

2.1 ANÁLISIS Y MAPEO DE CRÍMENES

En los últimos 20 años se han desarrollado muchas definiciones acerca de qué es el análisis de crímenes (*Crime Analysis*), dos de las cuales se mencionan en los siguientes párrafos.

El análisis de crímenes hace referencia a un conjunto de procesos sistemáticos y analíticos que proveen de información oportuna y pertinente acerca de patrones en los delitos y correlación de tendencias entre ellos. Diversos reportes y registros recolectan datos acerca de escenas del crimen, armas utilizadas, modus operandi, artículos robados, sospechosos, entre otros; el análisis y comparación de información entre un caso y otro entrega información valiosa que puede utilizarse en diversos ámbitos, permitiendo un mejor despliegue de recursos [7].

Se puede definir también como un conjunto de procesos dirigidos a proporcionar información de patrones y correlaciones para asistir a las labores administrativas y operacionales en la asignación de recursos para la prevención y eliminación de actividad criminal, asistir el proceso investigativo, aumentar la cantidad de aprehensiones y resolución de casos. Bajo este contexto, el análisis de crímenes asiste de muchas formas a las policías, como por ejemplo en el despliegue de patrullas, operaciones especiales, unidades tácticas, investigaciones, planificación, investigación, prevención de delitos, y servicios administrativos (presupuesto y planificación de programas).

Si bien existen diferencias en ambas definiciones, ambas tienen aspectos en común, como el hecho de asistir a las policías, el uso de información y métodos sistemáticos, y la entrega de información útil para una variedad de personas. En resumen, se puede decir que el análisis de crímenes es un campo de estudio y una práctica de la justicia criminal que utiliza información y métodos investigativos sistemáticos, para asistir a las policías y entregar información oculta tras los datos crudos [7].

Dentro del análisis de crímenes se pueden encontrar 3 niveles:

- i. **Nivel Táctico:** es el estudio de incidentes criminales recientes y de la actividad criminal potencial a través del estudio de características de cómo, cuándo, y dónde ha ocurrido actividad delictual, para asistir en el desarrollo de patrones, por ejemplo.
- ii. **Nivel Estratégico:** es el estudio de problemas criminales y otros asuntos relacionados con las policías para determinar patrones a largo plazo de la actividad delictual, y para evaluar las respuestas de las policías y procedimientos organizacionales.
- iii. **Nivel Administrativo:** es la presentación de los resultados acerca de la investigación de los delitos, y su análisis en base a asuntos legales, políticos y prácticos, para la transmisión de información a diferentes audiencias como agencias gubernamentales y la ciudadanía.

Una de las disciplinas de este campo de estudio corresponde al mapeo de crímenes (*Crime Mapping*), cuyo enfoque es la búsqueda de relaciones geográficas en la naturaleza misma del crimen y otras actividades delictuales. Su objetivo es conocer qué crimen está sucediendo en qué lugar, identificando los puntos donde ocurren la mayor cantidad de crímenes y analizando las relaciones espaciales entre ellos.

El mapeo de crímenes cumple tres funciones principales dentro del análisis de crímenes: facilitar los análisis estadísticos y visuales de la naturaleza del crimen en el espacio; vincular diferentes eventos en base a las variables geográficas y demográficas que tengan en común; asistir en la comunicación de los resultados del análisis realizado [8].



Figura 1: Relación entre análisis y mapeo de crímenes. El mapeo corresponde a una disciplina del análisis, pudiendo asistir en los tres niveles de éste: nivel táctico, estratégico y administrativo.

Dentro de los tres niveles del análisis de crímenes, el mapeo complementa cada una de ellas, de la forma que se describe a continuación [8]:

- i. **A nivel táctico:** utilizado para identificar patrones inmediatos (robo en áreas residenciales, robo en áreas comerciales, robo de autos, etc.)
- ii. **A nivel estratégico:** utilizado en aplicaciones a largo plazo, por ejemplo:
 - Para analizar la relación entre la actividad delictual y distintas variables demográficas (la tasa de desocupación, tasa de escolaridad, entre otras).
 - Para la asignación espacial y temporal de recursos (destinar oficiales o patrullajes en ciertas áreas más conflictivas).
 - Para examinar patrones cerca de localizaciones específicas, como escuelas o bares.
 - Para calcular la tasa de crímenes.
 - Para incorporar información cualitativa a los datos delictuales, como por ejemplo, lugares frecuentados por jóvenes, caminos frecuentes, o mercados de droga y prostitución.

- iii. **A nivel administrativo:** se transforma en una poderosa herramienta usada tanto por policías como medios de comunicación para transmitir la información al público.

2.2 MODELOS DE HOT SPOTS

Como se dijo anteriormente, una de las funciones principales del mapeo de crímenes es facilitar el análisis estadístico del crimen y su relación en el espacio. En la práctica, consiste en la detección de áreas con una gran densidad de delitos, y su análisis asiste a la policía en la identificación de áreas peligrosas, con el fin de conocer qué tipos de crímenes se cometen en qué lugar y la mejor forma de responder a ellos.

El crimen no distribuye de manera uniforme a lo largo de los mapas, y muy por el contrario, se acumula en algunas zonas mientras no se registra en otras [9]. Este conocimiento es utilizado por la gente, a veces sin saberlo, en las actividades diarias: evitando pasar por ciertas calles o barrios, por ejemplo, o refugiándose en sus casas en sectores más complicados. Incluso la decisión de dónde vivir, dónde comprar, o dónde estacionar el auto, se basan en el conocimiento popular de qué tan probable es ser víctima de algún delito en uno u otro lugar. Algunos piensan que esto puede ser solo respuesta del miedo que siente la gente, o que es una exageración en algunos casos, pero el hecho de que existan comportamientos diferentes en distintas zonas, sugiere que el crimen no distribuye de forma uniforme. La gente puede equivocarse en qué tan peligroso es un lugar, o tender a exagerar, pero no se equivoca al pensar que el riesgo de ser víctima de delitos es mayor en unos lugares que en otros, y la policía utiliza esta información diariamente. Aquellas áreas en las que el crimen se concentra, son denominadas “Hot Spots” [9].

Este término se ha utilizado de muchas maneras distintas, haciendo referencia a direcciones o lugares específicos, otros a cuadras conflictivas, o incluso a grupos de cuadras, pero todas poseen un enfoque similar: la búsqueda de una concentración de eventos individuales que puedan indicar una serie de crímenes relacionados; búsqueda de áreas pequeñas que tengan una gran cantidad de delitos; búsqueda de barrios o zonas de mucho conflicto para poder crear una conexión entre estos delitos y las condiciones sociales y demográficas. Como se puede apreciar, no existe una definición de uso común acerca de qué es un Hot Spot, pero se entiende generalmente como un

área que tiene un número de eventos delictuales mayor que el promedio, o un área donde las personas tienen un riesgo de victimización mayor que el promedio.

Ahora, si bien las definiciones de Hot Spot son similares, la diferencia que más se aprecia a simple vista es el área que abarca cada foco. Sin embargo, esta no es la diferencia de mayor relevancia: los factores que generan puntos peligrosos son diferentes de los que generan calles peligrosas, o vecindarios peligrosos, o incluso ciudades; y las medidas a seguir para lidiar con cada uno de ellos serán distintas.

Así, la teoría detrás del crimen es crítica para que el mapeo sea útil, pues permite la correcta interpretación de los datos y entrega una guía acerca de las acciones más apropiadas. Existen muchas teorías, y si bien tienen diferencias, no se contradicen unas con otras, sino más bien explican diferentes tipos de fenómenos criminales que ocurren en diferentes niveles geográficos. A continuación se explican las cuatro teorías más relevantes para este trabajo de título, representadas en la Figura 2 [9].

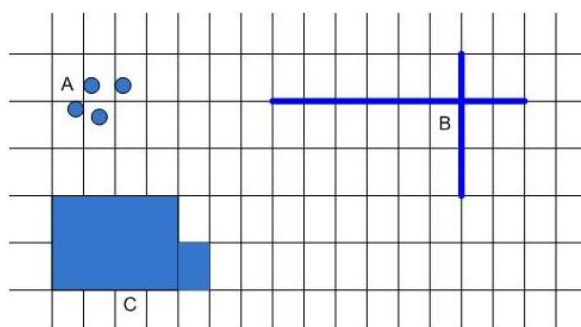


Figura 2: Diferencias entre las definiciones de Hot Spot. En A se aprecia teoría de lugares; en B la teoría de calles; en C la teoría de vecindarios.

- i. **Teoría de lugares (Place Theories):** explica por qué algunos eventos criminales ocurren en localizaciones específicas, trabajando entonces con los delitos que ocurren en el nivel más bajo de análisis. Involucra la observación de incidentes específicos y la formulación de preguntas del tipo: ¿en qué lugares ocurren robos y en qué lugares no ocurren? Los incidentes son entonces fenómenos puntuales, por lo que el análisis se basa en

información de direcciones, intersecciones de calles, u otras áreas pequeñas, típicamente representadas en un mapa por puntos. La acción policial es precisa pues se limita a direcciones exactas, como por ejemplo mandatos judiciales.

- ii. **Teoría de calles (Street Theories):** estudia los crímenes que ocurren en un nivel un poco más alto que puntos específicos, es decir, calles o manzanas. Intenta responder preguntas del tipo: ¿en qué calles se ejerce prostitución y en qué calles no? El análisis se basa entonces en información de calles, rutas, segmentos de autopistas, que serían en un mapa representados por líneas rectas o curvas. La acción policial ya no es tan precisa como el caso anterior, y principalmente consiste en patrullajes reiterados y concentración de esfuerzos por cambiar el tránsito por dichos lugares y por cambiar sus características.
- iii. **Teoría de vecindarios (Neighborhood Theories):** busca explicar las diferencias que existen entre vecindarios. Es un nivel mucho más alto que lugares específicos o calles, pues se trata de áreas mayores. Intenta entonces responder a preguntas del tipo: ¿qué áreas están afectadas por pandillas y qué áreas no? El análisis abarca información de manzanas, barrios, comunidades, o incluso de censos. Son típicamente representados por polígonos o elipses en el mapa, y la acción policial es mucho menos precisa pues las áreas son típicamente muy grandes para patrullajes reiterados. Sin embargo, dependiendo de las características del vecindario, la acción policial puede incluir esfuerzos para involucrar a los habitantes en la acción colectiva contra el crimen; si la delincuencia se mueve a lo largo de un área en vez de concentrarse en unos pocos lugares, entonces los esfuerzos deben ocurrir en este nivel.
- iv. **Otras teorías de áreas extensas:** existen otras teorías que intentan explicar los patrones criminales a niveles mucho mayores, por ejemplo, las teorías varían dependiendo de la ciudad y de la región. A nivel de ciudades, las acciones pueden incluir cambios económicos, de transporte, educacionales, de asistencia social, entre otras. A nivel regional, podría incluir un gran cambio social y político.

2.3 ACTIVIDAD POLICIAL PREDICTIVA

La Actividad Policial Predictiva (*Predictive Policing*) es la aplicación de técnicas analíticas que tienen como propósito identificar posibles objetivos para intervención policial, la prevención de

delitos, y la resolución de delitos pasados gracias a predicciones o pronósticos estadísticos. Si bien el uso de estadísticas y análisis espacial para predecir crimen han sido usados por décadas, solo hace unos pocos años surgió el interés por el uso de técnicas analíticas que se basan en conjuntos de datos de gran magnitud para hacer estas predicciones [10].

Las herramientas y las tecnologías de información utilizadas para este fin han sido desarrolladas y analizadas ampliamente por el mundo comercial, y muchas de las relaciones que existen en dicho sector pueden ser exploradas en el ámbito policial. Las policías usan información de delitos anteriores, el entorno en el que se producen, entre otros datos, para predecir y prevenir el crimen. La idea es tener mayor grado de conocimiento en los niveles tácticos y estratégicos, y desarrollar estrategias que fomenten una policía más eficiente y eficaz. Con la anticipación de la conducta humana, la policía puede identificar y desarrollar estrategias para prevenir la actividad criminal de delincuentes reincidentes contra víctimas reiteradas. Estos métodos también permiten a los departamentos de policía para trabajar de forma más proactiva con recursos limitados.

Una predicción en este sentido se trata como un pronóstico, es objetivo, científico y por sobre todo, repetible, por lo que su efectividad también debe poder traducirse a resultados tangibles, por ejemplo las tasas de delitos y de victimización deben disminuir, las tasas de arresto deben aumentar, etc.

Ahora, es natural preguntarse por qué el crimen es predecible. Existe una gran cantidad de evidencia que valida este hecho, básicamente porque las personas, incluidos los delincuentes, tienden a moverse dentro de las “Zonas de Confort”. Esto se traduce en que un delincuente tenderá a cometer un delito que haya cometido exitosamente en el pasado, generalmente cerca o en el mismo lugar. Si bien no es cierto en el total de los casos, ocurre con suficiente frecuencia para hacer que los métodos y relaciones estadísticas funcionen relativamente bien [10].

2.4 MODELOS PROBABILÍSTICOS PARA INCIDENTES CRIMINALES

Como se ha discutido antes, las fuerzas de seguridad monitorean los crímenes y delitos cometidos. Mediante la inclusión de información geográfica y demográfica, se buscan patrones tanto en tiempo como en espacio para predecir la actividad delictual futura. En otras palabras, se

tiene la necesidad de modelar patrones de riesgo asociados a la actividad delictual, pues con esto se pueden estudiar relaciones de causalidad entre ellos y estimar dónde y cuándo se producirá nuevamente un delito. Si esta predicción se realiza con precisión, es posible entonces desplegar más recursos para mejorar la seguridad, y para reducir las amenazas y el impacto de los delitos en la sociedad.

Muchas técnicas y modelos han sido desarrollados para poder cumplir este objetivo, siendo las más comunes las teorías de Hot Spots desarrolladas en el punto anterior. En ellas, los incidentes criminales se proyectan en un plano formando clusters o áreas de interés, y se asume que estos son constantes para cualquier horizonte de predicción. De esta forma, todas las predicciones futuras ocurrirán en las mismas áreas, puesto que estos modelos no consideran las relaciones existentes entre los delitos cometidos y el entorno. Así, los modelos de Hot Spots no reflejan cambios en los patrones de delitos a medida que el entorno en el que ocurren cambia. Por ejemplo, una calle puede tener un alto índice de delitos asociados debido a la poca iluminación de esta, pero si se instalasen nuevas luminarias, el patrón cambiaría.

Para evitar este problema se han desarrollado modelos estadísticos más sofisticados. Por ejemplo, en [11] se aplica un modelo en base a densidad de puntos, que depende de características asociadas a una localización. Estas características incluyen información geográfica, información demográfica, entre otras. En [12] y [13] se desarrolla un modelo en base a decisiones en el espacio: se asume que los delincuentes eligen lugares que puede modelarse como maximización de las utilidades, que a su vez depende tanto de la ganancia al cometer el delito como de la probabilidad de ser detenidos. Otros modelos pueden estudiarse en [14] y [15].

Sin embargo, ninguna de ellas incorpora directamente la componente temporal, y cuando se modela esta componente, por ejemplo usando series de tiempo, no se consideran las interacciones espacio-tiempo [16]. Estudios recientes han desarrollado algunos enfoques que utilizan modelos aditivos generalizados (*Generalized Additive Models, GAM*) para combinar información espacial, temporal y diversas características (por ejemplo, demográficas) para la predicción [17].

Matemáticamente, se tiene un área de interés $A \subset \mathfrak{R}^2$, un periodo de tiempo $T \subset \mathfrak{R}^+$, y un conjunto de servicios asociados a esta $\{S_{a,t} | a \in A, t \in T\}$. Para representar tanto el área como el tiempo se crea una partición de cada uno, en una cuadrícula $\{a_i\}$ para el área y en intervalos $\{t_j\}$ para el tiempo. En cada caso se cumple que la unión de las cuadrículas forma el área de interés, y la unión de los intervalos forma el periodo de tiempo. Además, $i, j \in \mathfrak{R}^+$ son índices. Los servicios asociados a una cuadrícula a_i y un intervalo t_j se representan por el vector S_{a_i, t_j} .

El objetivo es entonces la búsqueda de una función de probabilidad de que ocurra un incidente en un área y tiempo definidos, dados los servicios asociados a esa área y tiempo, es decir:

$$\mathcal{P}(\text{incidente}_{a_i, t_j} = 1 | S_{a_i, t_j})$$

Y la búsqueda de una función de decisión:

$$\delta_{a_i, t_j} = \delta \left[\mathcal{P}(\text{incidente}_{a_i, t_j} = 1 | S_{a_i, t_j}) \right]: [0,1] \rightarrow \{0,1\}$$

De tal forma que se minimice la función de pérdida dada por:

$$L = \frac{1}{N_A N_{T^*}} \cdot \sum_{a_i \in A, t_j \in T^*} \left[w_0 I(\delta_{a_i, t_j} = 0 | \text{incidente}_{a_i, t_j} = 1) + w_1 I(\delta_{a_i, t_j} = 1 | \text{incidente}_{a_i, t_j} = 0) \right] < \varepsilon$$

Aquí, $\text{incidente}_{a_i, t_j} = 1$ indica la ocurrencia de al menos un incidente o delito en el área a_i y tiempo t_j ; $\text{incidente}_{a_i, t_j} = 0$ indica que ningún incidente ocurre en dicha área y tiempo; $\mathcal{P}(\text{incidente}_{a_i, t_j} = 1)$ indica la probabilidad de que al menos un incidente ocurra en dicha área y período de tiempo; T^* representa un set de intervalos de tiempos futuros; N_A representa el número total de áreas o grillas; N_{T^*} representa el número total de intervalos de tiempo; $I(\cdot)$ es un indicador, w_0, w_1 representar los pesos de errores de diferente tipo; y finalmente ε es un umbral.

La mayor dificultad radica entonces en encontrar una función de distribución de probabilidad apropiada $\mathcal{P}(\text{incidente}_{a_i, t_j} = 1 | S_{a_i, t_j})$, de tal forma que tenga mayor verosimilitud en las zonas donde ocurren incidentes criminales, y menor verosimilitud en las zonas donde los

incidentes no ocurrirán en intervalos de tiempo futuros. Con esta función, las policías podrán formar su propia función de decisión en base a los recursos y preferencias de riesgo, por ejemplo.

Una adecuada caracterización de la función de distribución de probabilidad $\mathcal{P}(\text{incidente}_{a_i,t_j} = 1 \mid S_{a_i,t_j})$ [17] requiere de un esfuerzo computacional no menor y de la implementación integrada de una serie de métodos de análisis de datos.

En este Trabajo de Título se propone un enfoque mixto, que comienza con la creación de un modelo a priori utilizando información de servicios asociados a un área de interés y datos de incidentes criminales para aproximar una distribución de riesgo en torno a ellos, desarrollando así la componente espacial; la componente temporal se desarrolla mediante la inclusión sucesiva de información de nuevos delitos cometidos dentro de la misma área, variando la distribución inicial del riesgo en base a estas nuevas mediciones. Esto es equivalente a modelar la probabilidad de ocurrencia de un delito en un cierto lugar y tiempo, dados los servicios y eventos pasados asociados a ese lugar y tiempo. A continuación se describen en particular aquellas herramientas y conceptos que son necesarios para la implementación de la propuesta que representa el corazón de este Trabajo de Título.

2.5 MODELOS POR MEZCLA DE DISTRIBUCIONES (MIXTURE MODELS)

Hoy en día nos encontramos en una posición privilegiada para ser capaces de describir, estimar, predecir e inferir más detenidamente cualquier sistema de nuestro interés, gracias a métodos de cálculo cada vez más potentes y variadas distribuciones para modelar. Un ejemplo lo constituyen los llamados *Mixture Models*, pues dentro de una familia paramétrica ofrecen aproximaciones manejables en entornos no paramétricos; y basándose en distribuciones estándar permiten afrontar desafíos de alta complejidad computacional.

La definición de una mezcla de distribuciones es bastante sencilla: cualquier combinación convexa:

$$G(\vec{x}) = \sum_{i=1}^M w_i f_i(\vec{x}), \quad \sum_{i=1}^M w_i = 1 \quad M > 1$$

de cualquier distribución $f_i(\vec{x})$ con w_i el peso asociado a esa distribución de probabilidad, es una mezcla de distribuciones. En este Trabajo de Título se utiliza un caso particular en que las distribuciones f_i son gaussianas, denominándose así Modelos por Mezcla de Gaussianas (*Gaussian Mixture Models, GMM*) [18].

Así, GMM es una función de densidad de probabilidad paramétrica representada como una suma ponderada de las densidades de los componentes gaussianos, utilizadas comúnmente como un modelo paramétrico de la distribución de probabilidad de mediciones o características diversas. Considerando entonces la definición anterior, se tendrá que un GMM es una suma ponderada de M densidades, donde cada componente $f_i(\vec{x})$ es una Distribución Gaussiana multivariada (D dimensiones):

$$f_i(\vec{x}) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} \exp \left\{ -\frac{1}{2} (\vec{x} - \vec{\mu}_i)' \Sigma_i^{-1} (\vec{x} - \vec{\mu}_i) \right\}$$

Con vector de medias dado por $\vec{\mu}_i$ y matriz de covarianza dada por Σ_i . El GMM es parametrizado entonces por el vector de medias, las matrices de covarianza y los pesos de todas las densidades que lo componen [19]:

$$\lambda = \{w_i, \vec{\mu}_i, \Sigma_i\}_{i=1, \dots, M}$$

En la Figura 3 se puede apreciar una representación de un modelo por mezcla de gaussianas de M componentes.

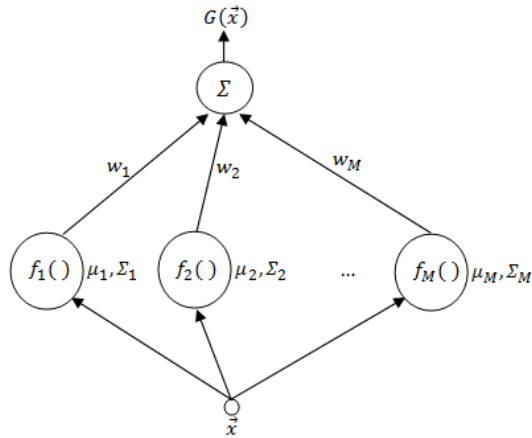


Figura 3: Representación de un GMM de M componentes. A partir de un vector de estados, se definen M distribuciones gaussianas, cada una con su vector de medias y matriz de covarianzas, y mediante una suma ponderada es posible obtener el modelo definitivo.

2.6 INFERENCIA BAYESIANA E INTEGRACIÓN DE MONTE CARLO

Diversas disciplinas científicas utilizan conjuntos de observaciones para estimar variables desconocidas. En el caso particular que se aborda en este Trabajo de Título, estas observaciones corresponden a incidentes criminales ocurridos en un área y tiempo determinados, y la variable a estimar es la distribución espacial de riesgo en dicha área.

Desde el punto de vista Bayesiano, es posible utilizar el conocimiento disponible del sistema en estudio para formular distribuciones a priori de la evolución de las cantidades desconocidas, y funciones de verosimilitud para relacionarlas con las observaciones. En este contexto, la inferencia sobre las variables del sistema se basa en la distribución a posteriori obtenida mediante el Teorema de Bayes.

Ahora bien, para realizar estimaciones en línea es necesario actualizar esta distribución a posteriori en forma secuencial cada vez que se reciben nuevas observaciones. Si las relaciones que definen los procesos de observación y transición de estado están dadas por un modelo de espacio-estado lineal y Gaussiano, es posible derivar una expresión analítica óptima para la evolución de las distribuciones a posteriori, ampliamente conocido como Filtro de Kalman [20].

Sin embargo, los procesos reales son bastante complejos como para ser descritos por este tipo de relaciones, y en consecuencia sucesión de densidades a posteriori a estimar son integrales multidimensionales que no tienen forma explícita para el caso general. Este problema es conocido por distintos nombres dependiendo del área en que se presente, entre los que se encuentran filtrado Bayesiano, filtrado óptimo no-lineal, filtrado estocástico, e inferencia en línea [21],[22],[23],[24].

Los métodos convencionales de Monte Carlo han sido extensamente usados para aproximar integrales y densidades de probabilidad. Sin embargo, su uso en el problema de inferencia Bayesiana no es directo, ya que éste involucra una secuencia de densidades de probabilidad variantes en el tiempo, mientras que los métodos convencionales asumen que la densidad objetivo es invariante. Por esta razón se hace necesario contar con una versión secuencial de los métodos de Monte Carlo, que permita incorporar las observaciones para mejorar las estimaciones realizadas recursivamente. Una herramienta que enfrenta este problema es Filtro de Partículas (FP), una técnica numérica para aproximar integrales multidimensionales con respecto a una secuencia de medidas de probabilidad [25]. Esta es una de las técnicas más utilizadas para resolver el problema del filtrado Bayesiano, principalmente por su facilidad de implementación y su amplio campo de aplicación. Estos métodos son capaces de aproximar una secuencia de medidas de probabilidad de dimensión creciente mediante un conjunto de muestras ponderadas del espacio de estado, las que evolucionan en base al conocimiento previo del sistema y a las observaciones obtenidas del proceso. La idea clave entonces es representar la densidad a posteriori por un conjunto de muestras aleatorias con pesos asociados, y calcular las estimaciones tomando como base estas muestras y pesos.

Bajo este contexto se define entonces el concepto de partícula, que corresponde a la dupla $\{x_{0:k}^i, w_k^i\}$, donde $x_{0:k}^i$ es una realización de la función de distribución de probabilidad del estado del proceso, a la que se le asocia un escalar w_k^i , versión muestreada de la función de distribución de probabilidad.

En este Trabajo de Título se propone una metodología basada en estos conceptos de inferencia Bayesiana y métodos secuenciales de Monte Carlo, por ejemplo la obtención de muestras (partículas) de una distribución a priori, la inclusión secuencial de nuevas observaciones y la

aproximación de una distribución a posteriori en base a estas. En la metodología propuesta, sin embargo, la información estadística está dada por la posición de las partículas y no por su peso, como sucede en los métodos genéricos de filtro de partículas. Así, solo se hace necesario estudiar la propagación y dispersión de las partículas a lo largo del tiempo para estimar distribuciones futuras aproximando un modelo GMM para cada una de ellas.

2.7 MUESTREO DE IMPORTANCIA (IMPORTANCE SAMPLING)

El principal problema al que uno se enfrenta al utilizar los métodos descritos en la Sección 2.6 es cómo obtener las muestras de la función de distribución de probabilidad. Si bien en este Trabajo de Título se trabaja con GMM (ver Sección 2.5), el uso de un modelo de tal complejidad para iteraciones sucesivas resultaría inviable, principalmente por la gran complejidad de cálculo. Es entonces relevante la búsqueda de algún método alternativo para la representación de estos modelos.

En estadística, el muestreo de importancia (*IS*) es una técnica general para la estimación de las propiedades de una distribución en particular utilizando muestras generadas a partir de una distribución diferente a la distribución de interés. La idea detrás de esto es utilizar un conjunto de muestras aleatorias y sus pesos respectivos, para aproximar la distribución deseada en base a ellos. Al considerar un número de muestras suficientemente grande, la aproximación calculada asemejará la distribución real.

Para explicar en detalle el procedimiento anterior, sea $\{x_{0:k}^i, w_k^i\}_{i=1}^{N_s}$ un conjunto de muestras aleatorias (partículas) que caracterizan una distribución a posteriori $p(x_{0:k} | z_{1:k})$, donde $\{z_{1:k}\}$ son observaciones del proceso, $\{x_{0:k}^i, i = 1, \dots, N_s\}$ son puntos de soporte con pesos asociados dados por $\{w_k^i, i = 1, \dots, N_s\}$, y en los cuales se tiene que $x_{0:k} = \{x_j, j = 1, \dots, k\}$ es un conjunto de puntos con todos los estados hasta el tiempo k . Por su parte los pesos se encuentran normalizados, de modo que $\sum_i w_k^i = 1$. Así, la distribución a posteriori puede aproximarse por [26]:

$$p(x_{0:k} | z_{1:k}) = \sum_{i=1}^{N_s} w_k^i \delta(x_{0:k} - x_{0:k}^i)$$

Se tiene entonces una aproximación discreta en base a pesos de cada muestra de la distribución real. Los pesos entonces son calculados en base al principio de IS, que se basa en el supuesto de que $p(x) \propto \pi(x)$ es una densidad de probabilidad de la cual es engorroso tomar muestras, pero $\pi(x)$ puede ser evaluada (y por ende para $p(x)$ también dada la proporcionalidad). Además, sean $x^i \sim q(x), i = 1, \dots, N_s$ muestras fácilmente generadas a partir de una función $q(x)$, llamada densidad de importancia. Entonces, una aproximación por pesos de la densidad $p(x)$ estará dada por:

$$p(x) \approx \sum_{i=1}^{N_s} w^i \delta(x - x^i)$$

Donde:

$$w^i \propto \frac{\pi(x)}{q(x)}$$

Entonces, si las muestras $x_{0:k}^i$ son obtenidas de una densidad de importancia $q(x_{0:k}|z_{1:k})$, entonces los pesos en la aproximación de la distribución a posteriori estarán dados por [26]:

$$w_k^i \propto \frac{p(x_{0:k}^i | z_{1:k})}{q(x_{0:k}^i | z_{1:k})}$$

2.8 REMUESTREO (RESAMPLING)

En estadística, se denomina como remuestreo o resampling a variados métodos que permiten:

- Estimar la precisión de muestras estadísticas mediante el uso de los datos disponibles (conocido como *jackknifing*).
- Tomar datos aleatoriamente con reemplazo a partir de un set de puntos (conocido como *bootsrapping*).
- Intercambiar marcadores de puntos de datos al realizar test de significancia.
- Validar modelos para el uso de subconjuntos aleatorios (validación cruzada).

Sin embargo, en el ámbito que se desarrolla en este trabajo de título, el remuestro es un método para reducir los efectos de la degeneración (problema que se presenta cuando solo algunas partículas tienen pesos relevantes, o dicho de otra forma, todas menos algunas tienen pesos despreciables) [26]. La idea básica entonces es eliminar las partículas que tienen poco peso y concentrarse en aquellas de alto peso, por cuanto están asociadas a regiones del espacio de estado con mayor probabilidad. Esto es generar un nuevo set de partículas $\{x_k^{i*}\}_{i=1}^{N_s}$ mediante un muestreo con reemplazo a partir de una aproximación discreta de una $p(x_k, z_{1:k})$ dada por:

$$p(x_k | z_{1:k}) = \sum_{i=1}^{N_s} w_k^i \delta(x_k - x_k^i)$$

De manera que $\mathcal{P}r(x_k^{i*} = x_k^j) = w_k^j$. La muestra resultante es entonces una muestra i.i.d. de una densidad discreta, por lo que los pesos se redefinen como $w_k^i = \frac{1}{N_s}$. Existen muchas formas de realizar este procedimiento, siendo la utilizada en este trabajo de título el algoritmo presentado en la Figura 4.

Algoritmo de Resampling

$$\left[\{x_k^{j*}, w_k^j, i^j\}_{j=1}^{N_s} \right] = \text{Resampling} \left[\{x_k^i, w_k^i\}_{i=1}^{N_s} \right]$$

- Inicializar CDF: $c_1 = 0$
- FOR $i = 2: N_s$
 Construir CDF: $c_i = c_{i-1} + w_k^i$
- END FOR
- Comenzar al principio de la CDF $i = 1$
- Seleccionar punto de inicio: $u_1 \sim \mathbb{U}[0, N_s^{-1}]$
- FOR $j = 1: N_s$
 Moverse a lo largo de la CDF: $u_j = u_1 + N_s^{-1}(j - 1)$
 WHILE $u_j > c_i$
 * $i = i + 1$
 END WHILE
 Asignar muestra: $x_k^{j*} = x_k^i$
 Asignar peso: $w_k^j = N_s^{-1}$
 Asignar padre: $i^j = i$
- END FOR

Figura 4: Operación del algoritmo de resampling utilizado.

2.9 FILTRO DE MÍNIMOS CUADRADOS RECURSIVO

Mínimos Cuadrados Recursivo (*Recursive Least Squares, RLS*) es un algoritmo que recursivamente encuentra coeficientes de un filtro que minimiza una función de costos de mínimos cuadrados ponderada en relación a la señal de entrada. En RLS se asume que la señal de entrada es determinística, a diferencia de otros algoritmos similares como en LMS (*Least Mean Squares*), que apuntan a minimizar el error cuadrático medio, donde la señal es estocástica. RLS muestra una convergencia bastante rápida, pero con un gran costo computacional asociado [27].

El propósito del filtro RLS es la minimización de una función de costo C mediante la selección apropiada de los coeficientes del filtro \mathbf{w}_n , actualizándolos a medida que nueva información se encuentra disponible, tal como se aprecia en el diagrama de la Figura 5. En ella se pueden apreciar las relaciones existentes entre la señal deseada, la señal estimada, el error de la estimación y señal recibida, junto con el esquema de actualización del algoritmo.

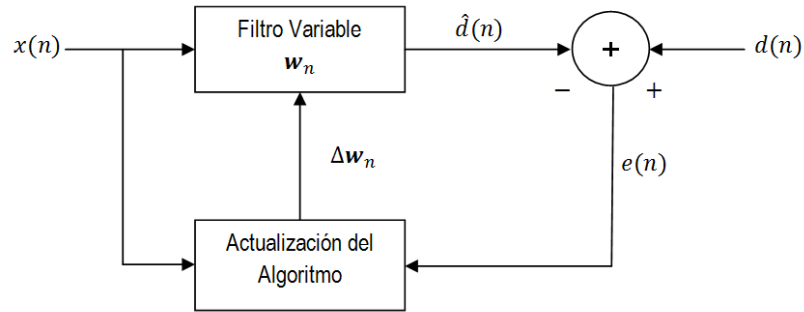


Figura 5: Esquema del algoritmo RLS. El error entre la señal deseada y la señal estimada por un filtro (variable) es utilizado para actualizar los parámetros del mismo, de forma recursiva cada vez que una nueva medición se encuentra disponible.

El algoritmo propuesto se basa en [28]. Se tiene entonces que el error depende de forma implícita de los coeficientes del filtro, pues depende de la estimación que este realiza de la señal deseada, es decir:

$$e(n) = d(n) - \hat{d}(n)$$

con la estimación de la señal deseada dada por: $\hat{d}(n) = \sum_{k=0}^p w_n(k)x(n-k) = \mathbf{w}_n^T \mathbf{x}_n$

Por su parte, la función de costos de mínimos cuadrados ponderada, aquella que se desea minimizar, al ser una función del error también depende de los parámetros del filtro, esto es:

$$C(\mathbf{w}_n) = \sum_{i=0}^n \lambda^{n-i} e^2(i)$$

donde λ es conocido como factor de olvido (*forgetting factor*), definido en el intervalo $[0,1]$, y cuya importancia radica en que exponencialmente asigna menor peso a muestras más antiguas. Mientras más pequeño sea su valor, menor será la contribución a la función de costo de las muestras previas, haciendo que el filtro sea más sensible a aquellas muestras más recientes, traducándose en mayores cambios en los coeficientes del filtro. En la práctica, si bien λ puede tomar valores en el intervalo $[0,1]$, los valores típicamente utilizados se encuentran en el intervalo $[0.95,1]$.

La función de costo se minimiza tomando las derivadas parciales para todas las entradas k (con $k = 0,1, \dots, p$) del vector de coeficientes del filtro \mathbf{w}_n e igualando a cero:

$$\frac{\partial C(\mathbf{w}_n)}{\partial w_n(k)} = \sum_{i=0}^n 2\lambda^{n-i} e(i) \frac{\partial e(i)}{\partial w_n(k)} = - \sum_{i=0}^n 2\lambda^{n-i} e(i) x(i-k) = 0$$

Reemplazando por la definición del error de estimación se tiene:

$$\sum_{i=0}^n 2\lambda^{n-i} [d(i) - \hat{d}(i)] x(i-k) = \sum_{i=0}^n 2\lambda^{n-i} \left[d(i) - \sum_{l=0}^p w_n(l) x(i-l) \right] x(i-k) = 0$$

Reordenando la igualdad anterior, es posible obtener la siguiente expresión:

$$\sum_{l=0}^p w_n(l) \left[\sum_{i=0}^n \lambda^{n-i} x(i-l) x(i-k) \right] = \sum_{i=0}^n \lambda^{n-i} d(i) x(i-k)$$

Esta expresión puede representarse como matrices y vectores de la forma:

$$\mathbf{R}_x(n) \mathbf{w}_n = \mathbf{r}_{dx}(n)$$

Donde $\mathbf{R}_x(n)$ es la matriz de covarianzas ponderada para $x(n)$; y $\mathbf{r}_{dx}(n)$ es la estimación equivalente a la covarianza cruzada entre $d(n)$ y $x(n)$. Finalmente, los coeficientes del filtro se encuentran despejando de la ecuación anterior:

$$\mathbf{w}_n = \mathbf{R}_x^{-1}(n)\mathbf{r}_{dx}(n)$$

Sin embargo, para aligerar la carga computacional se busca la obtención de una forma recursiva de estimación de los parámetros del filtro, es decir, una expresión de la forma:

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \Delta\mathbf{w}_{n-1}$$

Donde $\Delta\mathbf{w}_{n-1}$ es un factor de corrección. Se comienza por expresar la covarianza cruzada $\mathbf{r}_{dx}(n)$ en términos de $\mathbf{r}_{dx}(n-1)$, mediante la siguiente expresión:

$$\mathbf{r}_{dx}(n) = \lambda\mathbf{r}_{dx}(n-1) + d(n)\mathbf{x}(n)$$

Donde $\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-p)]^T$. Similarmente se expresa la matriz de covarianzas $\mathbf{R}_x(n)$ en términos de $\mathbf{R}_x(n-1)$ de la siguiente forma:

$$\mathbf{R}_x(n) = \lambda\mathbf{R}_x(n-1) + \mathbf{x}(n)\mathbf{x}^T(n)$$

Dado que para el cálculo de los coeficientes del filtro se requiere utilizar $\mathbf{R}_x^{-1}(n)$, gracias a la Matriz Identidad de Woodbury se tiene:

$$\mathbf{R}_x^{-1}(n) = \lambda^{-1}\mathbf{R}_x^{-1}(n-1) - \frac{\lambda^{-1}\mathbf{R}_x^{-1}(n-1)\mathbf{x}(n)\mathbf{x}^T(n)\lambda^{-1}\mathbf{R}_x^{-1}(n-1)}{1 + \mathbf{x}^T(n)\lambda^{-1}\mathbf{R}_x^{-1}(n-1)\mathbf{x}(n)}$$

Definiendo:

$$\mathbf{P}(n) = \mathbf{R}_x^{-1}(n) = \lambda^{-1}\mathbf{P}(n-1) - \mathbf{g}(n)\mathbf{x}^T(n)\lambda^{-1}\mathbf{P}(n-1)$$

Donde:

$$\mathbf{g}(n) = \frac{\lambda^{-1}\mathbf{P}(n-1)\mathbf{x}(n)}{1 + \mathbf{x}^T(n)\lambda^{-1}\mathbf{P}(n-1)\mathbf{x}(n)} = \frac{\mathbf{P}(n-1)\mathbf{x}(n)}{\lambda + \mathbf{x}^T(n)\mathbf{P}(n-1)\mathbf{x}(n)}$$

Modificando un poco la expresión, se tiene:

$$\mathbf{g}(n)\{1 + \mathbf{x}^T(n)\lambda^{-1}\mathbf{P}(n-1)\mathbf{x}(n)\} = \lambda^{-1}\mathbf{P}(n-1)\mathbf{x}(n)$$

$$\mathbf{g}(n) + \mathbf{g}(n)\mathbf{x}^T(n)\lambda^{-1}\mathbf{P}(n-1)\mathbf{x}(n) = \lambda^{-1}\mathbf{P}(n-1)\mathbf{x}(n)$$

$$\mathbf{g}(n) = \lambda^{-1}\mathbf{P}(n-1)\mathbf{x}(n) - \mathbf{g}(n)\mathbf{x}^T(n)\lambda^{-1}\mathbf{P}(n-1)\mathbf{x}(n)$$

$$\mathbf{g}(n) = \lambda^{-1}[\mathbf{P}(n-1) - \mathbf{g}(n)\mathbf{x}^T(n)\mathbf{P}(n-1)]\mathbf{x}(n)$$

Utilizando la expresión recursiva para $\mathbf{P}(n)$ se tiene finalmente:

$$\mathbf{g}(n) = \mathbf{P}(n)\mathbf{x}(n)$$

Volviendo a la recursión entonces, con estas nuevas definiciones, se tiene:

$$\mathbf{w}_n = \mathbf{P}(n)\mathbf{r}_{dx}(n) = \lambda\mathbf{P}(n)\mathbf{r}_{dx}(n-1) + d(n)\mathbf{P}(n)\mathbf{x}(n)$$

Reemplazando en esta ecuación la definición recursiva de la correlación cruzada $\mathbf{r}_{dx}(n)$, e incorporando la definición de $\mathbf{P}(n)$ y $\mathbf{g}(n)$, se puede obtener la expresión:

$$\mathbf{w}_n = \mathbf{P}(n-1)\mathbf{r}_{dx}(n-1) + \mathbf{g}(n)[d(n) - \mathbf{x}^T(n)\mathbf{P}(n-1)\mathbf{r}_{dx}(n-1)]$$

Ahora, considerando que $\mathbf{w}_{n-1} = \mathbf{P}(n-1)\mathbf{r}_{dx}(n-1)$ se llega a la ecuación de actualización dada por:

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \mathbf{g}(n)[d(n) - \mathbf{x}^T(n)\mathbf{w}_{n-1}] = \mathbf{w}_{n-1} + \mathbf{g}(n)\alpha(n)$$

Donde $\alpha(n) = d(n) - \mathbf{x}^T(n)\mathbf{w}_{n-1}$ es el error a priori. Comparándolo con el error a posteriori (el error calculado tras la actualización del filtro) se tiene:

$$e(n) = d(n) - \mathbf{x}^T(n)\mathbf{w}_n$$

Por lo que el factor de corrección finalmente será:

$$\Delta\mathbf{w}_{n-1} = \mathbf{g}(n)\alpha(n)$$

2.10 MÉTODO DE EVALUACIÓN DE MODELOS

En la Sección 2.4 se discute acerca de la función de pérdida que se desea minimizar y para lo cual se dispone de los modelos de probabilidad obtenidos con las técnicas anteriormente descritas [17]:

$$L = \frac{1}{N_A N_{T^*}} \cdot \sum_{a_i \in A, t_j \in T^*} \left[w_0 I(\delta_{a_i, t_j} = 0 | incidente_{a_i, t_j} = 1) + w_1 I(\delta_{a_i, t_j} = 1 | incidente_{a_i, t_j} = 0) \right]$$

$$L = \frac{1}{N_A N_{T^*}} \left\{ \sum_{a_i \in A, t_j \in T^*} w_0 I(\delta_{a_i, t_j} = 0 | incidente_{a_i, t_j} = 1) + \sum_{a_i \in A, t_j \in T^*} w_1 I(\delta_{a_i, t_j} = 1 | incidente_{a_i, t_j} = 0) \right\}$$

$$L = \frac{1}{N_S N_{T^*}} \{L_0 + L_1\}$$

La primera parte de esta función (L_0) es la suma ponderada de la cantidad de predicciones incorrectas para aquellas zonas donde los incidentes criminales efectivamente ocurren; la segunda parte (L_1) es la suma ponderada de predicciones incorrectas para las zonas donde no ocurren incidentes criminales. Para minimizar la función de costo descrita basta con minimizar ambas componentes. Para minimizar L_0 el modelo probabilístico debe predecir altas probabilidades en aquellas zonas donde los incidentes efectivamente ocurren. Por su parte, para minimizar L_1 las zonas de alta probabilidad deben ser escasas para un tiempo dado, debido a la escasez de incidentes criminales en toda el área de interés.

Ambos criterios descritos resultan de gran importancia en la caracterización de un modelo: el primer criterio significa que el modelo creado no debe dejar de predecir un foco en las zonas de alto riesgo, para que las policías puedan conocer gran parte de las localizaciones de incidentes futuros; el segundo criterio centra su importancia en la escasez de recursos, y el consiguiente problema de no poder patrullar muchas zonas a la vez. Basándose en estos criterios se propone el siguiente método para evaluar el desempeño de los modelos [29].

Para evaluar el desempeño en un tiempo dado t_j se calculan dos indicadores: HRP y TIP definidos respectivamente como:

$$HRP_\theta = \frac{\|\{a_i \mid \mathcal{P}(incidente_{a_i, t_j} = 1) > \theta\}\|}{\|\{a_i\}\|}$$

$$TIP_\theta = \frac{\|\{incidente_{a_i, t_j} = 1 \mid a_i \in \{a_i \mid \mathcal{P}(incidente_{a_i, t_j} = 1) > \theta\}\}\|}{\|\{incidente_{a_i, t_j} = 1\}\|}$$

HRP representa el porcentaje de áreas de alto riesgo predichas por el modelo, mientras que TIP representa el porcentaje de incidentes de un conjunto de prueba que ocurren dentro de estas áreas de alto riesgo. Por su parte, $\|\cdot\|$ representa el tamaño de un set, y θ representa un umbral definido en el intervalo $[0,1]$. Ambos indicadores, HRP y TIP, son calculados considerando diferentes valores para el umbral $\{\theta_i | \theta_i \in [0,1]\}$. Luego, TIP es graficado contra HRP. El gráfico resultante es similar a una curva ROC (*Receiver Operating Characteristics*): una técnica para visualizar, organizar y seleccionar clasificadores basados en su desempeño, y han sido utilizados en la teoría de detección de señales para representar el compromiso entre las tasas de éxito y tasas de falsas alarmas de clasificadores [30]. Bajo este concepto, se espera que la mayoría de los incidentes ocurran dentro de las áreas de alto riesgo (de un tamaño definido), y por ende un buen modelo presentaría una curva cercana a la esquina superior izquierda. Similar a los análisis de las curvas ROC, se utiliza como índice de comparación del desempeño de diferentes modelos el área bajo la curva AUC (*Area Under the Curve*): dado que un buen modelo presenta una curva cercana a la esquina superior izquierda, el AUC de un buen modelo debe ser cercano a la unidad, mientras que un modelo de estimación aleatoria tendrá un AUC=0.5, por lo que un mal modelo tendrá un AUC igual o inferior a este valor.

CAPÍTULO 3. MÉTODO DE GENERACIÓN AUTOMÁTICA DE DISTRIBUCIONES ESPACIALES DE RIESGO Y CARACTERIZACIÓN DE SU EVOLUCIÓN TEMPORAL

Dado que las policías requieren modelar patrones de riesgo asociados a la actividad delictual, se han creado muchos modelos para hacer frente tanto al estudio de relaciones de causalidad como a la estimación de lugar y tiempo de ocurrencia de un nuevo delito, buscándose una caracterización de la distribución espacial del riesgo. Sin embargo, en ellos se deja sin caracterizar, o caracterizada pobremente la evolución temporal del riesgo, por lo que su obsolescencia es segura.

El propósito de este capítulo es presentar un método empírico para la generación automática de distribuciones (modelos probabilísticos) espaciales de riesgo asociado a la actividad delictual, en conjunto con un mecanismo de caracterización de la evolución temporal de los mismos. El método propuesto se presenta esquemáticamente en la Figura 6.

Para cumplir con este propósito se divide el capítulo en tres secciones principales. Primero se describen los datos necesarios para la generación de los modelos. En segundo lugar se describe la generación del modelo, que dada la gran cantidad de datos disponible debe ser un proceso automático, para que la totalidad de ellos sea procesado bajo un criterio lógico común de forma eficiente. Para finalizar, se presenta un mecanismo para caracterizar la evolución temporal de la distribución de riesgo, mediante la inclusión iterativa y cronológica de nuevos eventos.

Se hace indispensable mencionar que ambas metodologías propuestas se basan en métodos empíricos, y algunos de los conceptos utilizados pueden requerir un mayor análisis para definir teóricamente la solución propuesta.

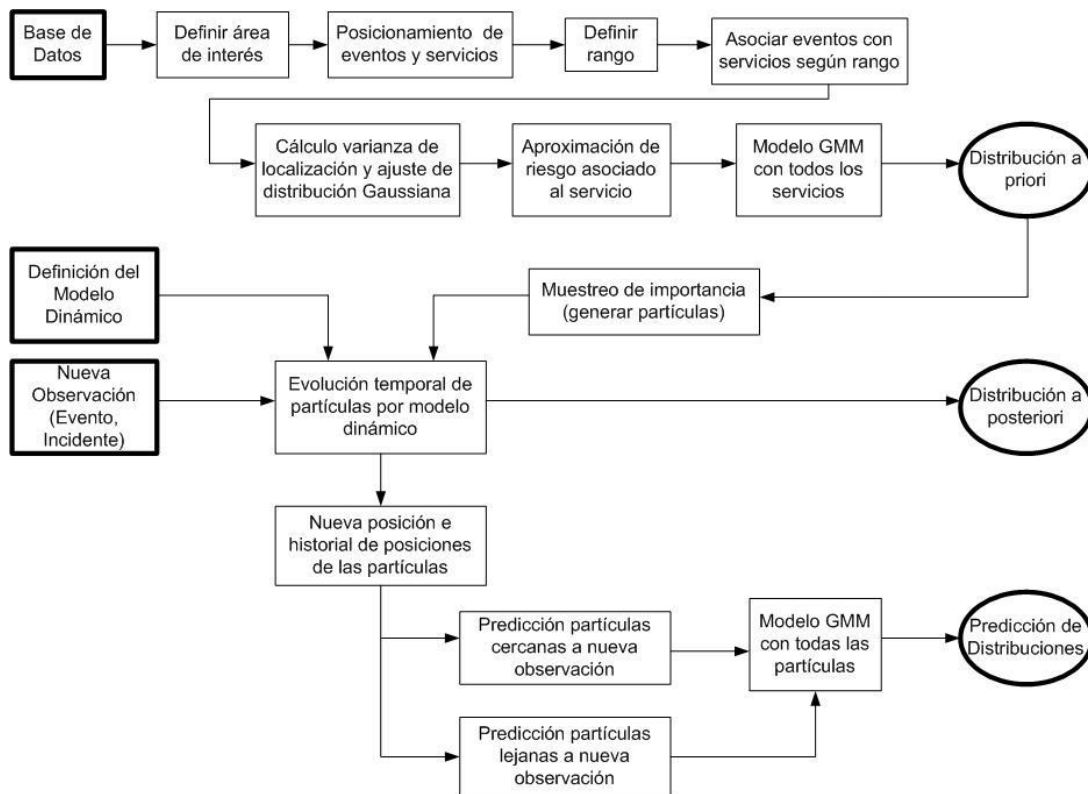


Figura 6: Esquema general del método propuesto para la generación de distribuciones espaciales de riesgo y caracterización de su evolución temporal.

3.1 INFORMACIÓN REQUERIDA

Para la creación de los modelos probabilísticos de riesgo asociado a incidentes delictuales es necesario contar con tres tipos de datos: definición del área de interés, datos georeferenciados de servicios asociados a dicha área, y finalmente datos georeferenciados de eventos delictuales.

3.1.1 DEFINICIÓN DEL ÁREA DE INTERÉS

Dado que se busca una distribución de probabilidad en el espacio, es de vital importancia definir el dominio de dicha función, o en otras palabras, definir el área $A \subseteq \mathcal{R}^2$ para la cual se busca el modelo probabilístico. Dicha área A dependerá de diversos factores, como por ejemplo, la razón por la que se genera el modelo, la capacidad de procesamiento de la máquina o la información disponible.

3.1.2 DATOS GEOREFERENCIADOS DE SERVICIOS ASOCIADOS AL ÁREA DE INTERÉS

En casi cualquier área de interés pueden encontrarse variados servicios (o *features*, en adelante s_i) ubicados dentro de su perímetro. Por servicios se entiende a diversos lugares característicos de cualquier asentamiento urbano, como hospitales, colegios, parques, supermercados, farmacias, entre otros, e incluso lugares geográficos como la presencia de ríos o lagos. Se asume de aquí en adelante que se busca un modelo para ciudades o zonas urbanas.

Los servicios que se consideran para un área A pueden variar enormemente dependiendo del tipo de modelo que se busque. Así, para un modelo puede incluirse desde lo más básico y frecuente, como bazares, hasta aquellos menos frecuentes, como municipios. O bien, para otro modelo podría solo considerarse solo un tipo de servicio si es que se buscara conocer el riesgo asociado únicamente a él, por ejemplo, para farmacias o cajeros automáticos. Sin embargo, es necesario mencionar que existe una relación entre el área de interés y la cantidad de servicios que se le asocian, principalmente porque si el área es muy grande y los servicios muy escasos, no se tendrá una representación apropiada del riesgo, excepto si se considera un rango demasiado grande (Figura 7, A). Por el contrario si el área es muy pequeña y los servicios demasiados, no se mejorará necesariamente el modelo pero si se sobrecargará el procesamiento (Figura 7, B).

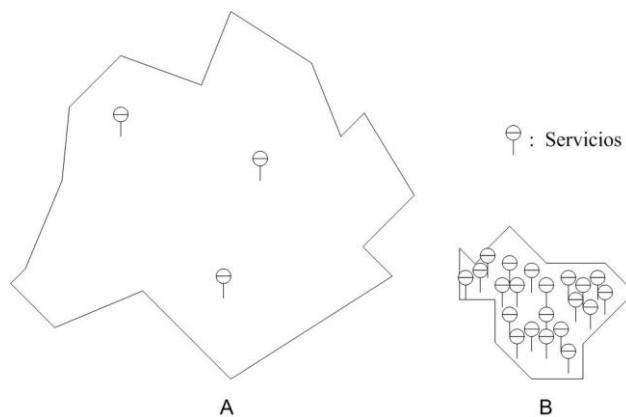


Figura 7: Representación de la relación entre el área de interés y la cantidad de servicios asociados. Se aprecia que el uso de pocos servicios no representará fielmente todas las áreas de riesgo, y el uso de muchos servicios solo produce un aumento en el gasto computacional.

3.1.3 DATOS GEOREFERENCIADOS DE EVENTOS DELICTUALES ASOCIADOS AL ÁREA DE INTERÉS

Por eventos delictuales se entiende a cualquier delito (en adelante d_j) que haya ocurrido en un intervalo de tiempo (en adelante $T \subseteq \mathfrak{R}^+$) y del que se tenga registro. De la cantidad de delitos considerados en el modelo dependerá su validez estadística, por lo que es necesario contar con un número considerable de eventos. Nuevamente esta “cantidad considerable” dependerá del área de interés: para áreas mayores se requerirá una mayor cantidad de datos.

El tipo de delitos para el cual se genera el modelo puede variar: pueden generarse modelos para uno o más tipos de delitos, dependiendo de si se quiere un modelo general de riesgo para una zona en particular, o bien con un grupo de delitos, por ejemplo los delitos de mayor connotación social (DMCS).

3.2 GENERACIÓN DE MODELOS ESPACIALES DE RIESGO

Debido a la necesidad de contar con una herramienta de generación automática de patrones de riesgo, que pueda ser utilizada en la predicción de delitos, se presenta en esta sección una metodología de generación automática de modelos probabilísticos de riesgo. El objetivo de esta metodología es la creación, a partir de datos de servicios y eventos delictuales, de un modelo probabilístico que caracterice el riesgo asociado a la actividad delictual, y que pueda ser utilizado para la predicción de los focos de criminalidad de un área determinada. Esta metodología tiene como eje principal la asociación de eventos delictuales con la presencia de servicios, que sirven para tener representatividad de toda el área de interés a medida que la cantidad de servicios aumenta. Los eventos, más específicamente la varianza en el espacio de sus localizaciones, permiten aproximar una distribución normal bivariada en cada servicio, y mediante la generación de un modelo por mezcla de gaussianas (GMM) se genera la caracterización probabilística del riesgo final.

Con la finalidad de entregar una descripción completa del esquema implementado, se divide esta sección en dos partes. Primero se presenta una descripción general de lo que el algoritmo realiza en cada etapa, permitiendo al lector comprender a grandes rasgos la implementación. Segundo, se

presenta detalladamente cada uno de los pasos necesarios para cumplir con el objetivo mismo de la metodología.

3.2.1 DESCRIPCIÓN GENERAL DEL ESQUEMA

El esquema propuesto se puede fraccionar en las siguientes etapas, presentadas en la Figura 8.

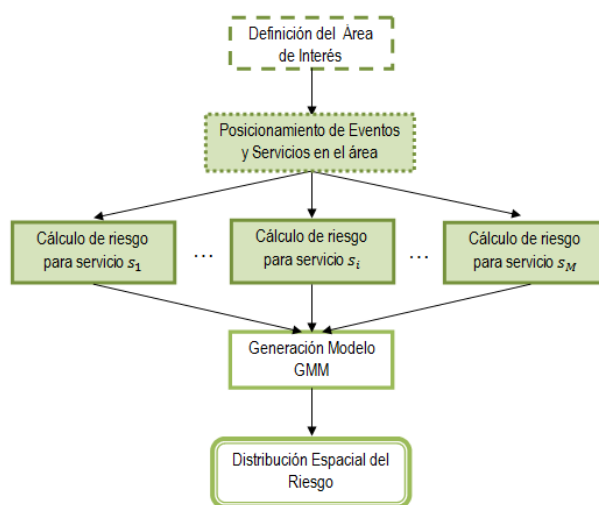


Figura 8: Diagrama de flujo general de la metodología de generación automática de modelos de riesgo.

Como es posible observar en la Figura 8, la metodología propuesta parte con la definición del área de interés, pues en función de esta se especificarán tanto los servicios presentes (y a considerar en el modelo) como los delitos con los que se trabajará en el resto del esquema. Al considerar entonces un área dada, se definen necesariamente los servicios y eventos, por lo que se espera que estos sean representativos del caso de estudio. Esto tiene como implicancias la cantidad y variedad de servicios que se consideran para la generación del modelo, pues como se explicó en la Sección 3.1.2, una caracterización insuficiente del área llevará a un modelo poco representativo de la realidad, y una caracterización recargada llevará a encarecer (computacionalmente hablando) el procesamiento. Aun así, esta etapa de generación de modelo se lleva a cabo solo una vez, por lo que un procesamiento pesado no afecta la implementación en línea propuesta para la caracterización de la evolución temporal.

Una vez definida el área de interés (y por ende servicios y eventos) se procede a la creación de un mapa, en el cual se posicionan los eventos gracias a la información georeferenciada que se tiene disponible de cada uno de ellos (Figura 9, elemento A). Se realiza el mismo procedimiento para cada servicio de forma separada (Figura 9, elemento B). Se define para cada uno de ellos un rango, que a priori es idéntico para todos. Aquí por rango se entiende a una distancia radial que toma como centro la localización del servicio, y a cada uno se le asocian los eventos cuya distancia a este sea inferior al rango (Figura 9, elemento C).

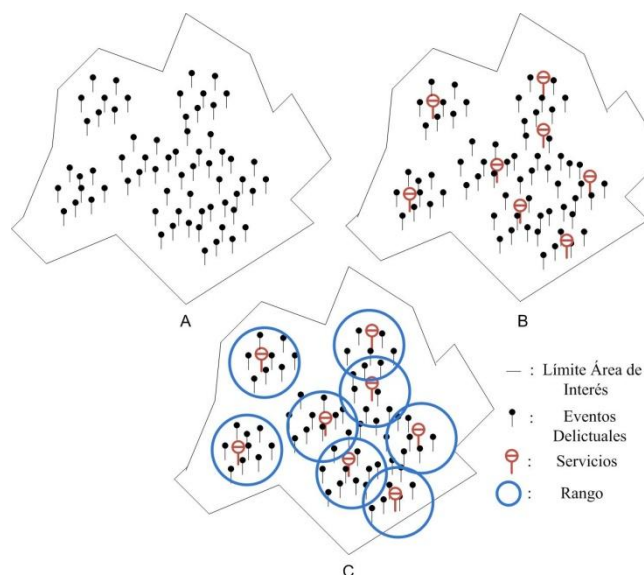


Figura 9: Representación de la localización de eventos (A), eventos y servicios (B), y eventos, servicios y el rango para cada uno de ellos (C). Solo corresponde a una representación y muy probablemente no todos los eventos delictuales se ubiquen cercanos a los servicios.

Se procede entonces a la etapa denominada “Cálculo del riesgo para servicio s_i ”. Mediante la asociación de eventos a cada servicio, es posible calcular la varianza en la localización de los delitos, medida que se utiliza para aproximar una distribución bivariada centrada en el servicio. Se dijo anteriormente que “a priori” el rango es idéntico para todos los servicios, pero existen casos específicos es que el rango debe aumentarse para poder hacer la aproximación de la distribución (ver Sección 3.2.2 para más detalles). Al final de esta etapa, para cada servicio se tendrá una distribución de riesgo aproximada por una distribución normal bivariada, cuyo promedio es la localización del servicio, y cuya matriz de varianza está dada por la varianza empírica de la localización de los eventos delictuales que ocurren dentro de un rango a su alrededor.

Finalmente, gracias a que por cada uno de los servicios se tiene una distribución del riesgo que es utilizada en la creación de un GMM, se obtiene un modelo generalizado para toda el área de interés, que representa una función de distribución de probabilidad a priori del riesgo en el espacio.

3.2.2 DESCRIPCIÓN DETALLADA DEL ESQUEMA

Para realizar una explicación en detalle, se considera el diagrama de la Figura 10. En las siguientes líneas se explicará en profundidad cada uno de los bloques de la metodología propuesta.

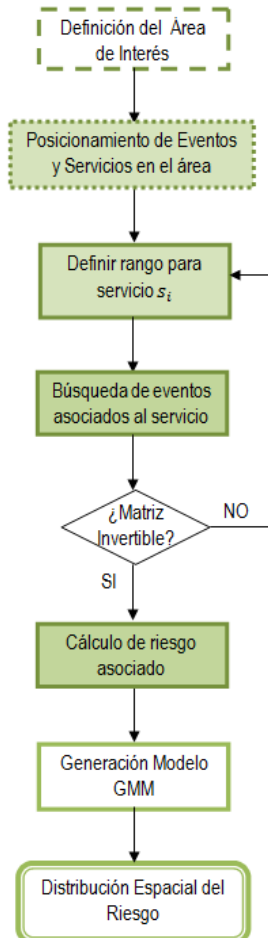


Figura 10: Diagrama de flujo detallado de la metodología de generación automática de modelos de riesgo.

3.2.2.1 Definición del Área de Interés

Sea un área de interés $A \subseteq \mathfrak{R}^2$, definida según el propósito de generación del modelo. No requiere mayor detalle que el presentado en la Sección 3.2.1

3.2.2.2 Posicionamiento de Eventos y Servicios en el área

Sea un conjunto de eventos delictuales definidos por $\{d_j\}_{j=1}^D$ ocurridos en un intervalo de tiempo $T \subseteq \mathfrak{R}^+$; sea también un conjunto de servicios $\{s_i\}_{i=1}^M$ presentes en A durante el intervalo de tiempo T . De ambos conjuntos se tiene información de su posicionamiento, puesto que la creación de un mapa y su localización en él solo dependerá de las coordenadas que se quiera utilizar (preferentemente coordenadas GPS). No requiere mayor detalle que el presentado en la Sección 3.2.1

3.2.2.3 Definir rango para el servicios s_i

Sea $r_i \in \mathfrak{R}^+$ el rango asociado a un servicio s_i , tal que r_i define el radio de una circunferencia, cuyo centro son las coordenadas del servicio que corresponde. A priori, es idéntico para todos los servicios, es decir $\{r_i\}_{i=1}^M = R$. Sin embargo, como se explica a continuación, existen excepciones.

3.2.2.4 Búsqueda de eventos asociados a un servicio

Aquí se considera un servicio s_i fijo, por lo que el procedimiento descrito a continuación se repite para cada uno de ellos.

Sea entonces la localización de un delito d_j dada por el par ordenado $\vec{x}_d = (d_{j_x}, d_{j_y}) \in \mathfrak{R}^2$, y sea también la localización de un servicio s_i dada por el par ordenado $\vec{x}_s = (s_{i_x}, s_{i_y}) \in \mathfrak{R}^2$. Se calcula para cada uno de los delitos:

$$dist_{ij} = \sqrt{(d_{j_x} - s_{i_x})^2 + (d_{j_y} - s_{i_y})^2}$$

Si $dist_{ij} \leq r_i$, se “asocia” el delito d_j al servicio s_i . Repitiendo para todos los delitos de $\{d_j\}_{j=1}^D$, finalmente se obtendrá un nuevo conjunto de eventos que están asociados a un servicio en particular, defínase como $\mathfrak{D}_i = \{\vec{x}_d; \text{tales que } dist_{ij} \leq r_i\}$.

3.2.2.5 Condicionante y cálculo del riesgo asociado

Utilizando todos los elementos de \mathfrak{D}_i se calcula la varianza de las localizaciones, denótese por $\Sigma_{\mathfrak{D}_i} \in \mathcal{M}_{2 \times 2}$. Esta matriz de varianzas será utilizada para aproximar una distribución normal bivariada f_i , cuyo centro sea la localización del servicio s_i :

$$f_i(\vec{x}) = \frac{1}{2\pi|\Sigma_{\mathfrak{D}_i}|^{1/2}} \exp\left\{-\frac{1}{2}(\vec{x} - \vec{\mu}_{\mathfrak{D}_i})' \Sigma_{\mathfrak{D}_i}^{-1}(\vec{x} - \vec{\mu}_{\mathfrak{D}_i})\right\}$$

Donde $\Sigma_{\mathfrak{D}_i}$ es la matriz de varianzas; $\vec{\mu}_{\mathfrak{D}_i} = \vec{x}_s$ el vector de medias. De aquí entonces resulta natural la razón para redefinir en algunos casos el rango de un servicio: cuando muy pocos datos cumplen la condición $dist_{ij} \leq r_i$ (servicio ubicado en una zona con baja tasa de victimización), o bien cuando muchos datos se repiten (por ejemplo, un servicio que ha sufrido victimización reiterada) la matriz de varianzas es cercana a cero y por ende no es invertible. En estos casos se procede a aumentar el rango, y por ende a agregar más datos al conjunto \mathfrak{D}_i hasta que la matriz de varianzas sea invertible. Así, a la salida de este bloque se tendrá una distribución de riesgo de incidentes delictuales asociada a cada servicio.

3.2.2.6 Generación de modelo por GMM

Dado que para cada servicio se tendrá una distribución de riesgo, en esta etapa solo basta con generar el modelo por mezcla de gaussianas, esto es, calcular:

$$PDF_{a\ priori}(\vec{x}) = \sum_{i=1}^M w_i f_i(\vec{x})$$

Donde:

$$w_i = \frac{1}{M} \Rightarrow \sum_{i=1}^M w_i = 1$$

Entonces, al final de este método, se tiene un GMM, que en el fondo es una distribución de probabilidad, que caracteriza espacialmente el riesgo de un área de interés.

En la Figura 11 se observa un ejemplo de cómo es el modelo obtenido al finalizar el método de generación automática de modelos de riesgo. Se aprecia en dicha figura las curvas de nivel para hipotéticos servicios y eventos asociados a cada uno de ellos.

En la Sección 3.3, y por ende en la segunda parte de esta metodología, se trabaja a partir de la distribución de probabilidad (en adelante distribución a priori) entregada al finalizar la etapa de generación de modelo, por lo que la caracterización del riesgo que se obtenga es de vital importancia para el correcto funcionamiento del resto del procedimiento.

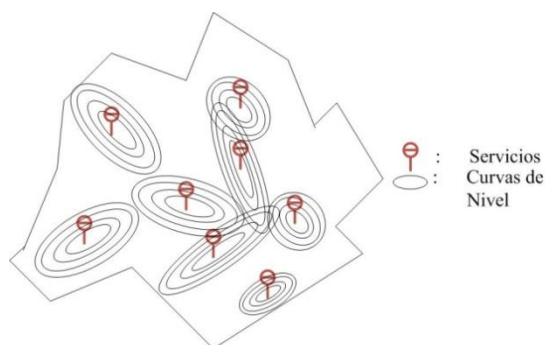


Figura 11: Ejemplo de las curvas de nivel de un modelo hipotético obtenido al final del método propuesto. Cada una de las distribuciones gaussianas se encuentra centrada en la posición del servicio que generó la zona de riesgo, orientándose automáticamente según la distribución original de los incidentes, y no de acuerdo al rango considerado (es decir, no todas las distribuciones serán iguales).

3.3 CARACTERIZACIÓN DE LA EVOLUCIÓN TEMPORAL DE LA DISTRIBUCIÓN ESPACIAL DE RIESGO

Si bien la caracterización espacial es necesaria para poder cumplir con el objetivo de predicción, no es suficiente por sí sola. Se requiere la inclusión de variables temporales para una caracterización completa, que se adapte a cambios en las condiciones del medio que generan la delincuencia. El objetivo de esta etapa es la caracterización de la evolución temporal de la distribución espacial de riesgo asociada a toda un área de interés mediante una metodología inspirada en los métodos secuenciales de Monte Carlo, la definición de un modelo dinámico y un

filtro RLS, incluyendo cronológica y sucesivamente la información de nuevos delitos, obteniéndose un esquema de cambio en línea.

Para lograr esto, se toma como punto de partida la distribución a priori y un muestreo de importancia de esta, y cada una de las partículas que aproxima la distribución a priori es modificada por un modelo dinámico común. Luego, dependiendo del movimiento generado y de patrones obtenidos por observaciones previas, se predice una nueva posición para la partícula. Dado que todos los pesos son iguales, para aproximar una distribución a posteriori se utiliza un modelo por mezcla de gaussianas, que permite volver a tener una distribución en lugar de muestras aisladas. Si bien para un número suficientemente grande de partículas la distribución y la aproximación por partículas es equivalente, se requiere volver al espacio continuo para una evaluación del modelo generado.

Para que el lector comprenda a cabalidad el esquema propuesto, esta sección se ha dividido en dos partes. En primer lugar se describe de forma general cada una de las etapas del método, permitiendo comprender a gran escala el propósito de esta etapa. Posteriormente, se presenta una descripción detallada de cada etapa, con lo que se logra un entendimiento profundo de cómo se incluye la variable temporal al modelo, que si bien no es explícita, permite obtener una evolución de la distribución del riesgo a medida que nuevas observaciones (nuevos delitos o incidentes delictuales) ocurren.

3.3.1 DESCRIPCIÓN GENERAL DEL ESQUEMA

La metodología propuesta se puede dividir en dos grandes etapas: offline y online. Sin embargo, existe una división aún más importante, y tiene relación al objetivo que cumple cada uno de los bloques presentados en la Figura 12. Bajo este concepto, se pueden notar entonces dos etapas: el análisis secuencial de los datos (todo lo realizado hasta el bloque de Evolución Temporal) y la predicción propiamente tal (Módulo de Predicción). En primera instancia se requiere estudiar un conjunto de datos utilizado de forma secuencial, para comprender las tendencias que sigue la distribución espacial de riesgo. Solo teniendo estas tendencias es posible continuar a la etapa siguiente, que corresponde a la predicción a corto y largo plazo.

Observando la Figura 12 es posible notar que esta metodología toma como entrada al proceso la salida del proceso anterior (ver Sección 3.2), correspondiente a una distribución de probabilidad definida en toda el área de interés A , y nuevas observaciones de delitos que se incorporan a medida que ocurren (cronológicamente). Esta etapa es iterativa, y se lleva a cabo cada vez que se registra un nuevo evento delictual, modificando así secuencialmente la distribución a priori para encontrar una nueva distribución que añada la última información disponible y se caracterice así una distribución a posteriori variante en el tiempo.

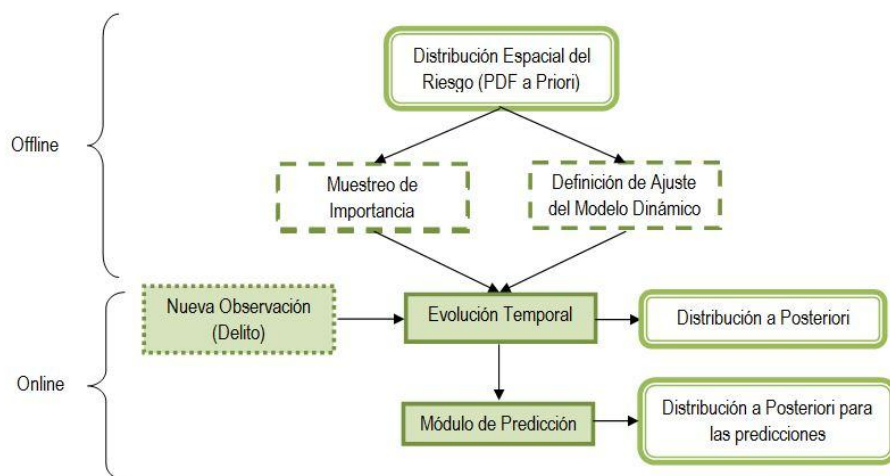


Figura 12: Diagrama de flujo general de la metodología de caracterización de la evolución temporal.

La metodología comienza con la selección de muestras representativas de la función de densidad de probabilidad del riesgo (la distribución a priori), mediante un algoritmo de muestreo de importancia dada la dificultad de muestrear a partir de la misma función. El objetivo de este bloque es encontrar un set de partículas (muestras) que aproximen la distribución de riesgo, para simplificar el manejo (ver Sección 2.7).

Teniendo ya una nueva observación y un set de partículas definido tras el muestreo de importancia, para cada una de las partículas se calcula la distancia entre esta y la observación actual. Dependiendo de esta distancia se determina si la posición futura de la partícula será afectada por el ruido del proceso únicamente o por la observación y el ruido, según un modelo dinámico responsable de la evolución temporal de las partículas y por ende del modelo de probabilidad. En la Figura 13 se puede apreciar un ejemplo de esta asociación: dependiendo de la

distancia, las partículas serán afectadas por el ruido o la observación, como se explica en el siguiente párrafo.

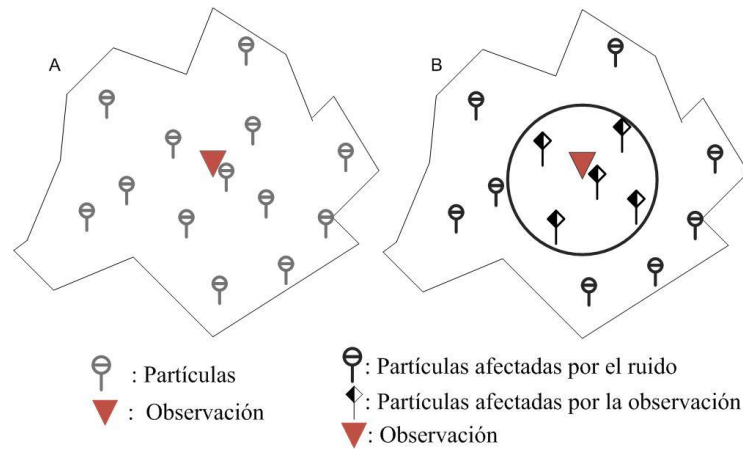


Figura 13: Ejemplo del umbral de operación del modelo dinámico. Del total de partículas en un área de interés dada (A), tras la definición de un umbral de operación solo algunas de ellas serán afectadas por la observación (B).

Se propone un modelo dinámico para cada partícula, en el que el movimiento dependa de la distancia entre la partícula y la observación. La idea de este modelo es modificar la posición de las partículas mediante solo el ruido cuando se encuentran muy lejos o muy cerca de la observación actual, y generar un acercamiento gradual de las partículas que se encuentren dentro de cierto umbral. Así, algunas partículas oscilarán cerca de una posición dada, al ser afectadas solo por el ruido, pero aquellas que sean afectadas por la observación (ubicadas dentro de un umbral) tendrán una tendencia a acercarse a ésta: esta tendencia puede modelarse mediante Mínimos Cuadrados Recursivos (RLS, descrito en la Sección 2.9). En este punto se cumple entonces la primera gran etapa: el análisis secuencial de datos.

En la Figura 14 se aprecia el efecto descrito: partículas que se encuentran alejadas de la observación solo serán afectadas por el ruido, manteniéndose lejanas; mientras que las partículas que se encuentran a cierta distancia de ella tenderán a acercarse.

Teniendo ya información acerca de las tendencias que sigue la distribución del riesgo, es posible comenzar la segunda gran etapa: la predicción del riesgo a posteriori. Con la predicción mediante RLS de la posición de una partícula condicional a la última observación adquirida, se genera una zona de riesgo asociada a ella, que tiene directa relación con el ruido de proceso. Recordemos

que cada partícula es una representante de la distribución a priori del riesgo, y su posición en el mapa será un indicador de riesgo en esa zona, pues la posición es afectada por la presencia de nuevas observaciones. Sin embargo es una variable aleatoria y no un valor puntual como lo sería la partícula en específico, por lo que una distribución de densidad de probabilidad se asocia a cada una de ellas, con el cuidado de caracterizar la incertidumbre correspondiente (explicado en detalle en la Sección 3.3.2).

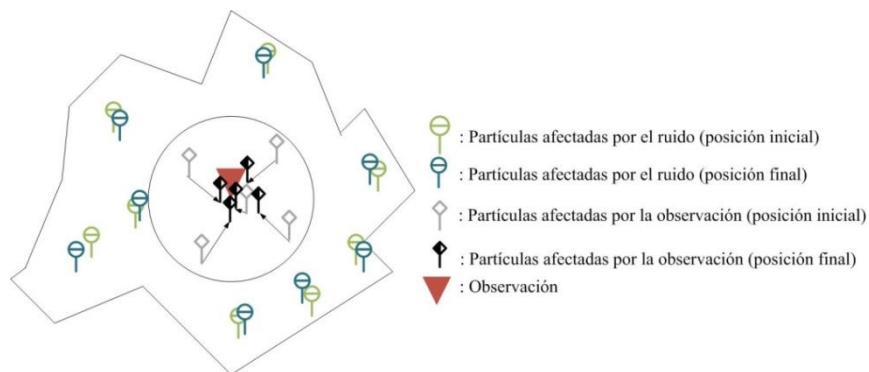


Figura 14: Ejemplo del movimiento de las partículas por efecto de la observación y del modelo dinámico. Se aprecia que aquellas partículas lejanas (fuera del umbral de operación) se desplazan solo por efecto del ruido (en verde y azul), mientras que aquellas ubicadas dentro del umbral de operación experimentarán un desplazamiento mayor generado por la atracción que genera la observación (gris y negro).

Finalmente, y al realizar este procedimiento para todas las partículas, se puede aproximar un nuevo modelo por mezcla de Gaussianas para volver a una distribución espacial de riesgo, pero modificada por la presencia de una nueva observación. Reiteradas modificaciones hacen posible la caracterización de la evolución temporal del modelo puesto que cualquier cambio en las condiciones del entorno se verá reflejado en las localizaciones de los nuevos eventos, que son utilizadas para modificar la distribución a priori hasta una distribución representativa de los nuevos eventos y por ende de los cambios en el entorno.

3.3.2 DESCRIPCIÓN DETALLADA DEL ESQUEMA

Para realizar una explicación detallada, se considera en primera instancia el diagrama presentado en la Figura 15. En ella se presentan los bloques catalogados como “Offline” en la Figura 12, solo por motivos didácticos para una mejor explicación de los procedimientos utilizados en esta etapa.

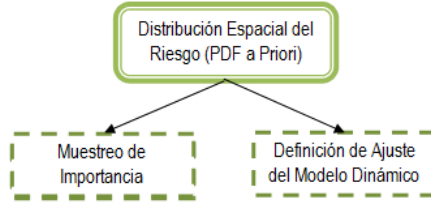


Figura 15: Diagrama de flujo detallado de la metodología offline de caracterización de la evolución temporal y predicciones del riesgo espacial.

3.3.2.1 Muestreo de Importancia

Sea una distribución de probabilidad, dada en primera instancia por la PDF a priori encontrada en la primera parte de esta metodología (ver Sección 3.2), que de acuerdo a lo expresado en [referencia a capítulo 2, muestreo de importancia], puede ser expresada de la siguiente forma:

$$p(x) \approx \sum_{i=1}^{N_s} w^i \delta(x - x^i)$$

Donde los pesos w^i se calculan a través de:

$$w^i \propto \frac{\pi(x)}{q(x)}$$

En este caso, las muestras $x^i \sim q(x), i = 1, \dots, N_s$ son generadas fácilmente a partir de la densidad de importancia $q(x)$, que en este caso particular se define como una distribución uniforme cuyo dominio es el área de interés $A \subseteq \mathfrak{R}^2$. Para la generación de las muestras se definen dos variables aleatorias con distribuciones uniformes definidas en latitud (definida entre coordenadas a, b) y longitud (definida entre coordenadas c, d) respectivamente, de la forma:

$$X_1 \sim \mathbb{U}_{[a,b]}$$

$$X_2 \sim \mathbb{U}_{[c,d]}$$

Donde $\mathbb{U}_{[a,b]}$ representa una distribución uniforme en el intervalo $[a, b]$ con los límites incluidos. Al crear pares ordenados de muestras obtenidas de X_1, X_2 , se obtienen las muestras requeridas $x^i \sim q(x), i = 1, \dots, N_s$, donde N_s es un parámetro de diseño que representa la cantidad de

muestras con la cual se pretende aproximar la distribución (por ende, mientras mayor sea el número mejor será la aproximación, pero también será más caro computacionalmente).

Una vez generadas las muestras, estas estarán ubicadas uniformemente en el plano, con los pesos asociados proporcionales a la PDF a priori. Sin embargo se busca una aproximación de esta distribución y dado que cantidad de partículas generadas no puede ser tan elevada como para sobrecargar el procesamiento, se requiere la realización de *resampling*. Como se explica en la Sección 2.8, esto permite la eliminación de las partículas de bajo peso, y la consecuente creación de partículas de alto peso para tener una mayor concentración de estas en las zonas más probables y aproximar de mejor manera la distribución. Es necesario añadir un ruido proporcional al ruido de proceso para otorgar una mayor variabilidad al proceso, pero manteniendo a la vez la importancia que se le da a una zona en particular.

En la Figura 16 se puede apreciar un simple ejemplo de lo descrito anteriormente. En ella se aprecia en Elemento A la generación de muestras a partir de la densidad de importancia, que distribuye uniforme en toda el área A, delimitada por el rectángulo exterior. Mediante el algoritmo de *resampling* apreciado en el Elemento B, es posible obtener un nuevo set de partículas concentradas en las zonas de mayor probabilidad, pero teniendo aun representantes de las zonas menos probables.

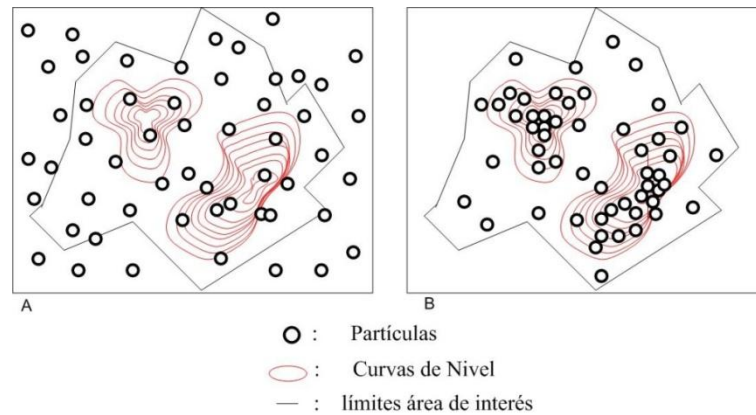


Figura 16: Ejemplo de muestreo de importancia y resampling. En A se aprecian las partículas obtenidas mediante el muestreo de importancia, que distribuye uniforme en el área de interés. En B se aprecian las partículas obtenidas después del algoritmo de resampling, presentando una mayor densidad en las zonas de mayor probabilidad de acuerdo a la distribución a priori.

Así, al finalizar este bloque, se dispone de un conjunto de partículas con una posición y un peso propio para aproximar la distribución. El resto del algoritmo se basa en la modificación de la posición de cada partícula y no en la modificación de los pesos, por lo que la aproximación dependerá de la concentración de partículas en una zona y no del peso de la partícula en ella.

3.3.2.2 Definición del ajuste del modelo dinámico

Como fue mencionado, el método propuesto se basa en la posición de las partículas, más específicamente en su movimiento a medida que se agregan nuevas observaciones, por lo que es necesario definir un modelo variante en el tiempo para desplazarlas en el plano. En esta sección se explica tanto el modelo dinámico propuesto como el ajuste necesario para su correcto funcionamiento.

El modelo dinámico debe depender de tres variables: la posición de la partícula en el instante anterior, la observación actual, y un ruido de proceso. Además, debe cumplir con tres requisitos: las partículas lejanas a la observación deben mantenerse alejadas y ser afectadas solo por el ruido de proceso; las partículas cercanas a la observación deben mantenerse cerca, afectadas solo por el ruido de proceso; y por último, las partículas a una distancia prudente de la observación (parámetro de diseño) deben acercarse a ella, y ser afectadas por el ruido de proceso. Se define entonces:

$$x(k) = x(k - 1) + \beta(d)\{y(k) - x(k - 1)\} + w(k)$$

Donde:

- La posición de la partícula se denota $x(k)$ y es la posición de la partícula al ser afectada por la observación.
- La posición en el instante de tiempo anterior se denota $x(k - 1)$.
- La observación que en este instante de tiempo afecta a las partículas es $y(k)$.
- El ruido de proceso está dado por $w(k)$ definido de modo que su varianza sean tres cuabras (parámetro de diseño) y media nula.
- La expresión $\{y(k) - x(k - 1)\}$ entrega una dirección de movimiento y su magnitud (recordar que ambos son pares ordenados).

- La función $f(d)$ es una función dependiente de la diferencia d entre la observación y la partícula que este siendo estudiada (recordar que este proceso se lleva a cabo para todas las partículas separadamente). La definición de esta función es aquella que produce mayores cambios al modelo, y de su ajuste depende en gran medida el funcionamiento del modelo dinámico propuesto.

Para definir completamente el modelo dinámico basta entonces con encontrar la función $f(d)$, que permita el cumplimiento de las condiciones propuestas. Esta función se define entonces asemejando una distribución Gaussiana bivariada, pero ajustada de tal forma que el máximo de esta sea la unidad. Es decir:

$$g(d) = \frac{1}{2\pi|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\vec{d} - \vec{y})' \Sigma^{-1}(\vec{d} - \vec{y})\right\}$$

$$f(d) = \frac{g(d)}{\max(g(d))}$$

Para esto, la matriz diagonal Σ es un parámetro de diseño y depende del área de interés, definiendo un rango dentro del cual las observaciones comienzan a moverse hacia la observación. En esta etapa debe definirse este parámetro, que se mantendrá constante durante toda la etapa “online”. A grandes rasgos, puede sostenerse que los componentes de esta matriz son directamente proporcionales al área A : mayor el área de interés, mayores serán los coeficientes de la matriz de covarianza; menor área de interés, menores serán los coeficientes de la matriz de covarianza. Dependen sin embargo de cuánto se desea que una sola observación afecte el mapa completo, por lo que los valores finales son de criterio del usuario. A partir de la definición de Σ , se define un umbral de acción que indica el límite aproximado para el cual el efecto de la observación deja de ser notorio en los desplazamientos de las partículas.

Es necesario notar que bajo este esquema:

- Partículas lejanas: a medida que la diferencia se hace mayor, los valores que toma la función son cada vez más pequeños, y a cierta distancia se pueden aproximar a cero. Luego: $f(d)\{y(k) - x(k - 1)\} \approx 0$.

- Partículas cercanas: a medida que la diferencia se hace más pequeña, la función se acerca a su valor máximo. En el límite, la diferencia es nula y la función toma valor unitario. Así, $f(d)\{y(k) - x(k - 1)\} \approx 0$.
- Partículas intermedias: $f(d)\{y(k) - x(k - 1)\}$ tiene un valor acotado, que permite el acercamiento gradual de las partículas.

Entonces, para definir empíricamente el modelo es necesario realizar pruebas modificando la matriz de covarianzas hasta obtener el ajuste deseado en las posiciones de las partículas. Se define una matriz a priori y se estudia el efecto en el movimiento: si el movimiento es excesivo, se redefine una matriz con coeficientes menores a los anteriores; por el contrario si el movimiento es demasiado leve, se redefinen los coeficientes a valores mayores. Así, al finalizar este estudio previo, se conocen todos los parámetros del modelo dinámico para su posterior uso en bloques siguientes.

Ahora, en la Figura 17 se muestra el esquema detallado de la implementación online de la caracterización. Esta etapa se lleva a cabo cada vez que una nueva observación se registra. En los siguientes párrafos se explica detalladamente cada uno de los bloques que conforman esta etapa.

3.3.2.3 Modelo dinámico para todas las partículas

Teniéndose definidos ya desde la etapa “offline” los parámetros del modelo dinámico, en esta etapa sólo se hace necesario aplicar este modelo a cada una de las partículas. Tal como se explicó anteriormente, para cada partícula se tendrá una nueva posición dependiente de la observación actual dada por la expresión:

$$x(k) = x(k - 1) + f(d)\{y(k) - x(k - 1)\} + w(k)$$

Recordar que la función de ajuste $f(d)$ ya se encuentra definida de la etapa anterior, y que $w(k)$ es el ruido de proceso, considerado un parámetro de diseño. Una vez realizado el ajuste de las posiciones para cada partícula $p_i \in \{p_i\}_{i=1}^{N_S}$, es posible encontrar una distribución espacial del riesgo a posteriori mediante la generación de un modelo por mezcla de gaussianas, tal como se explica en la Sección 2.5; esta distribución entonces incluye la información de la nueva observación.

Como no solo se requiere una actualización del modelo sino también predicciones para tiempos futuros, es necesario definir un esquema de predicción del movimiento de una partícula (y por ende de la zona de riesgo que representa) en base a la información proporcionada por la nueva observación y el movimiento ya provocado. Los bloques presentados en la Figura 17 se llevan a cabo para cada partícula $p_i \in \{p_i\}_{i=1}^{N_s}$ con el fin de obtener estas predicciones en el movimiento y con ello en la distribución de densidad de probabilidad que se aproxima a partir de las posiciones de las partículas.

3.3.2.4 Decisión de metodología a seguir

Si bien todas las partículas son modificadas por el modelo dinámico, no todas son modificadas de igual manera. Se divide entonces todo el set de partículas en dos conjuntos: aquellas afectadas por la observación mayoritariamente y aquellas afectadas por el ruido de proceso mayoritariamente, de la siguiente forma:

$$d_{op} = \sqrt{(x_o - x_{p_i})^2 + (y_o - y_{p_i})^2}$$

$$\mathcal{A} = \{p_i \mid d_{op} \leq u\}$$

$$\mathcal{N} = \{p_i \mid d_{op} > u\}$$

Donde d_{op} es la distancia entre la observación actual y la partícula en estudio; u es el umbral de acción de la función de ajuste del modelo dinámico; \mathcal{A} es el conjunto de partículas afectadas por la observación mayoritariamente y \mathcal{N} el conjunto de partículas afectadas por el ruido de proceso mayoritariamente. En otras palabras, si se define:

$$x(k) = x(k-1) + f(d)\{y(k) - x(k-1)\} + w(k)$$

De esta forma, para las partículas afectadas por la observación mayoritariamente se cumple $f(d)\{y(k) - x(k-1)\} \gg w(k)$, mientras que para las partículas afectadas por el ruido mayoritariamente se cumple $f(d)\{y(k) - x(k-1)\} \ll w(k)$.

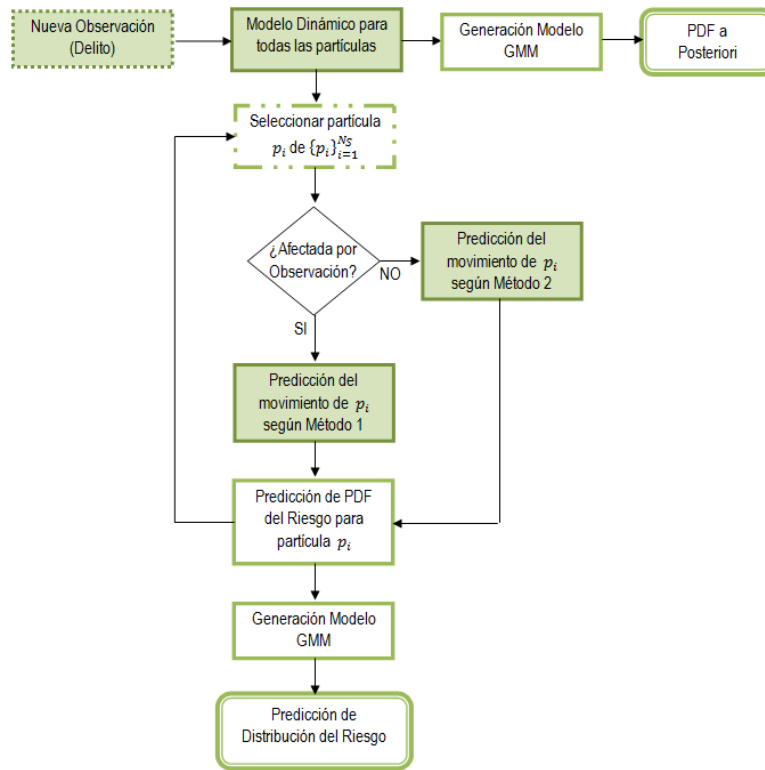


Figura 17: Diagrama de flujo detallado de la metodología online de caracterización de la evolución temporal y predicciones.

A partir de este momento, las partículas seguirán metodologías diferentes dependiendo del conjunto al cual pertenezca. Es necesario mencionar que esta división puede omitirse y utilizar para todas las partículas el mismo criterio. Sin embargo esto significa un gasto computacional innecesario, por lo que se omite en los casos en que no es relevante.

3.3.2.5 Análisis de tendencias y predicción del movimiento de las partículas

Se ha decidido unificar estos dos pasos para efectos de mayor comprensión por parte del lector, siendo que son claramente diferentes. Sin embargo se cree que, dado que se presentan dos métodos con marcadas diferencias entre sí, es preferible presentar el análisis y la predicción juntos, para no perderse en la lectura.

En esta etapa se implementan dos métodos, dependiendo del conjunto al cual pertenezca la partícula estudiada: si $p_i \in \mathcal{A}$ se ejecuta el Método 1; si $p_i \in \mathcal{N}$ se ejecuta el Método 2.

Ambas metodologías aquí presentadas son netamente empíricas, y por ende algunos conceptos, como por ejemplo la convergencia de la distribución condicional que se aborda más adelante, requieren de mayor análisis para determinar aspectos teóricos. Sin embargo, las pruebas realizadas y presentadas en el capítulo siguiente son suficientes para proponerlas como una solución factible al problema planteado.

En la Figura 18 se muestran los diagramas de flujo para el Método 1 en A, y Método 2 en B. El Método 1 es relativamente más complejo que el Método 2, principalmente porque éste incluye el filtro RLS. Esta división se realiza para aligerar el costo computacional, pero en caso de disponer de mayor poder de cómputo es perfectamente factible tratar a todas las partículas de la misma forma y utilizar indistintamente el Método 1.

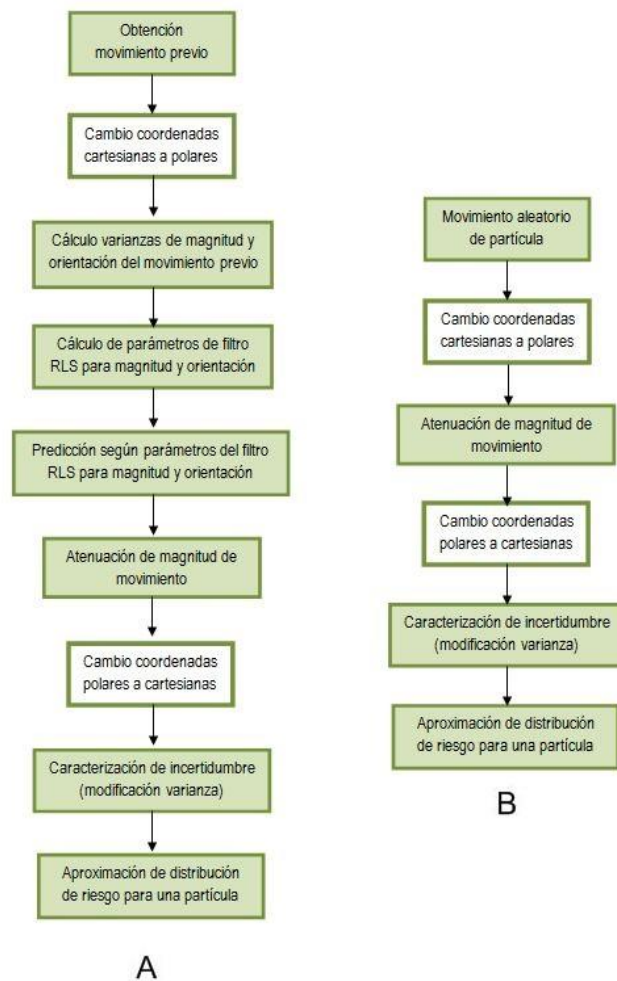


Figura 18: Diagramas de flujo de los Métodos 1 y 2, implementados para la predicción.

i. Método 1: Partículas afectadas por la observación mayoritariamente

Si la posición de la partícula fue efectivamente modificada por la observación, el movimiento de la partícula tendrá cierta tendencia, que puede fácilmente transformarse de coordenadas cartesianas a coordenadas polares, permitiendo una representación en magnitud (ρ) y dirección (θ) del movimiento.

Como se puede apreciar en el ejemplo de la Figura 19, una partícula que fue afectada en varias iteraciones (o solo una inclusive) por la presencia de la observación (A) mostrará una tendencia tanto en módulo (B) como en orientación (C) hacia donde ésta se encuentra. Los gráficos presentados son solo referenciales, mostrando solo la tendencia que describe la partícula y no la magnitud y orientación exacta.

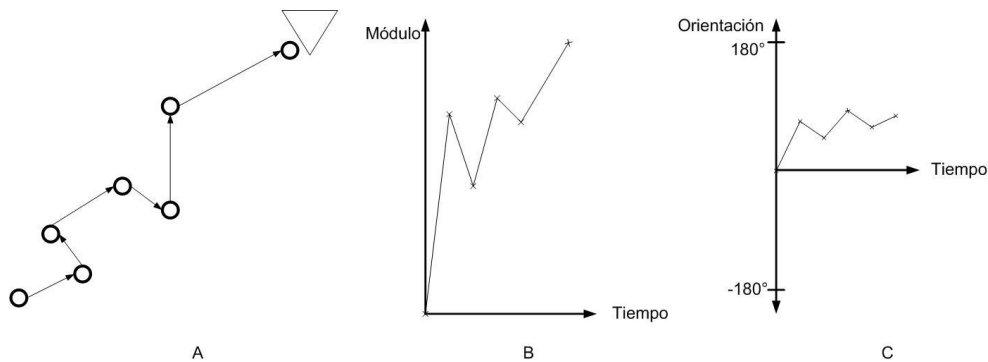


Figura 19: Ejemplo de la caracterización del movimiento de una partícula afectada por la observación mayoritariamente. En A se presenta el desplazamiento real que experimenta una partícula dadas sucesivas observaciones; estos desplazamientos secuenciales se expresan en coordenadas polares, presentándose el módulo o magnitud del movimiento y el ángulo u orientación, en B y C respectivamente.

Entonces, para obtener una predicción de la posición futura de una partícula, se comienza con la caracterización descrita, transformando el movimiento de coordenadas cartesianas a coordenadas polares, tomando como referencia la posición inicial de la partícula (del muestreo de importancia y resampling en un comienzo, y del inicio de una ventana de tiempo tras algunas iteraciones). Con esta caracterización del movimiento se calcula la varianza de cada parámetro, σ_ρ^2 y σ_θ^2 para la magnitud y la orientación respectivamente.

Mediante RLS se obtiene una ecuación de actualización para la magnitud y la orientación (dos filtros RLS separados) utilizando para este fin la inclusión de un ruido de media cero y varianza σ_p^2 o σ_θ^2 según corresponda. Se tiene entonces una ecuación de actualización de la forma:

$$z(t) = \varphi_1 z(t-1) + \varphi_2 z(t-2) + \dots + \varphi_p z(t-p)$$

Donde $z(t)$ es el parámetro que se quiere estimar (en este caso magnitud y orientación), y los parámetros del filtro $\varphi_1, \varphi_2, \dots, \varphi_p$ se obtienen mediante RLS. Con los parámetros de este, es posible entonces obtener una predicción tanto para la magnitud como para la orientación que posteriormente se expresan en coordenadas cartesianas.

Ahora bien, este procedimiento se repite cuantas veces sea necesario para obtener la cantidad de predicciones requeridas, que se define como la cantidad de mediciones regresivas que utiliza RLS para la estimación de parámetros. Es decir, si para estimar $z(t)$ se utilizan mediciones hasta $z(t-p)$, las predicciones a futuro se realizarán hasta $z(t+p)$, para ser consistentes con la confiabilidad del modelo.

En la Figura 20 se aprecia un esquema de la metodología implementada. En A se presenta el movimiento de una partícula en específico durante algunas iteraciones, expresado en el plano cartesiano. Este movimiento se presenta nuevamente en B pero esta vez en coordenadas polares, tomando como referencia (ρ_0, θ_0) . Mediante filtro RLS para la magnitud y la orientación es posible obtener una predicción para estas variables (ρ^*, θ^*) , que se representa nuevamente en coordenadas polares (C y D respectivamente).

Existen sin embargo dos pequeñas modificaciones a considerar para predicciones a dos o más pasos, cuya finalidad es la atenuación de la magnitud del movimiento (un concepto de inercia que “frene” a las partículas para no desvirtuar el modelo con movimientos prolongados) y el aumento de la incertidumbre a medida que se predice a una mayor cantidad de pasos. Ambos criterios resultan de vital importancia para obtener un modelo y predicciones razonables.

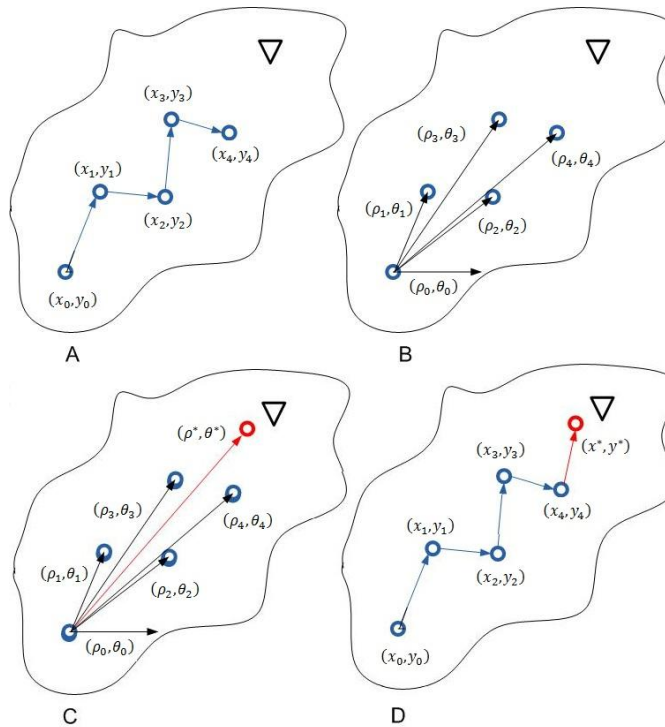


Figura 20: Esquema de las transformaciones y geometría del Método 1. En A se muestra el desplazamiento total que ha experimentado una partícula dada, en coordenadas cartesianas, y luego expresado en coordenadas polares en B. Mediante dos filtros RLS es posible obtener una predicción para la magnitud y la orientación del próximo movimiento presentado en C, y esta nueva posición predicha se transforma al plano cartesiano nuevamente en D.

El primer criterio debe su importancia a que el movimiento de las partículas hacia la observación más reciente se realiza con una finalidad exploratoria, para poder enriquecer el modelo con la nueva información que esa localización específica puede entregarle. Si se sostiene en el tiempo el mismo desplazamiento esta finalidad se pierde, pues la localización de la partícula se aleja de la zona en la que efectivamente ocurrió el nuevo evento, perdiéndose el concepto de la obtención de una distribución de densidad de probabilidad dadas las últimas observaciones disponibles. Es decir, se produce un cambio notorio en la distribución sin la presencia de una observación que le indique dicho cambio. El segundo criterio resulta algo más natural de explicar, puesto que a medida que uno más se aleja de la última observación, del último dato certero del que se dispone, la incertidumbre acerca de la verdadera posición trascurrido el movimiento se hace mayor.

Así, para predicciones a dos o más pasos, se proponen las siguientes recursiones para la magnitud del movimiento y la varianza de predicción:

$$\rho(t) = \lambda^{t-1} \rho(t-1) \quad ; \quad t \geq 2$$

$$\sigma_{pred}^2(t) = \frac{\sigma_{pred}^2(t-1)}{\lambda^{t-1}} ; t \geq 2$$

Donde λ es el factor de olvido utilizado en el filtro RLS, y las condiciones iniciales se definen como:

$$\rho(1): RLS$$

$$\sigma_{pred}^2(1): Varianza de Proceso$$

En la Figura 21 se aprecia el primer criterio referente a la atenuación de la magnitud de movimiento. En A se puede apreciar la primera predicción, realizada puramente mediante RLS y denotada (ρ^*, θ^*) , y una segunda predicción también mediante RLS denotada (ρ^{**}, θ^{**}) . Atenuando la predicción de la magnitud mediante la fórmula descrita, es posible obtener un nuevo punto denotado $(\rho_{pred}, \theta^{**})$. En B se muestra esta misma situación pero de vuelta al plano cartesiano.

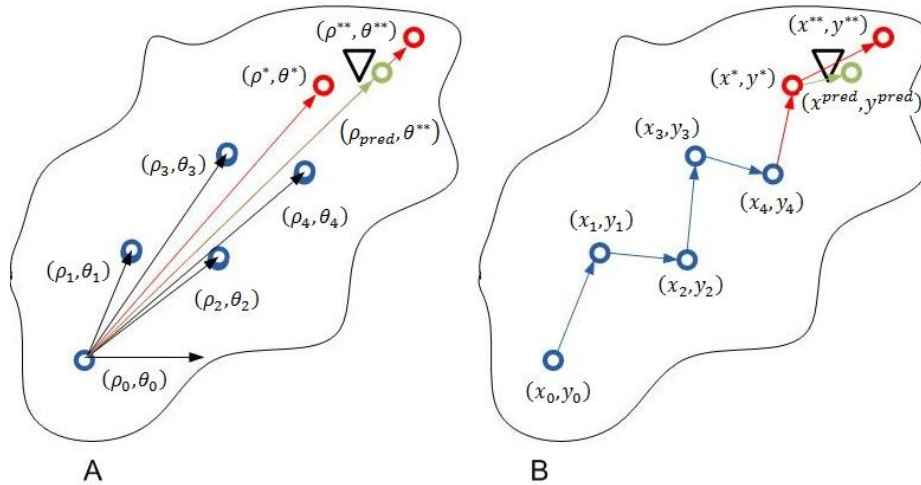


Figura 21: Representación de criterio de atenuación para el Método 1. En rojo se presentan los valores reales que se estiman para la posición futura de la partícula, tanto en el plano polar (A) como cartesiano (B). En verde se presentan los valores atenuados para la posición futura de la partícula, tanto en el plano polar (A) como cartesiano (B).

Así, al finalizar este proceso, se tendrá una distribución de la densidad de probabilidad del riesgo asociado a una partícula en específico para las predicciones a p pasos.

ii. Método 2: Partículas afectadas por el ruido mayoritariamente

Si una partícula se encuentra fuera del umbral de acción, y por ende el efecto que posee la observación es menor al efecto producido por el ruido de proceso, las posiciones futuras de dicha partícula tendrán un comportamiento aleatorio, pues nada incita un cambio en su comportamiento al estar lejos de la última observación.

Así, el método de predicción solo busca desplazar a la partícula en base al ruido de proceso, con las mismas consideraciones del Método 1: atenuación de la magnitud del movimiento y ajuste de la varianza a medida que aumenta el horizonte de predicción.

El Método 2 comienza entonces por la definición de un ruido aditivo de media cero y varianza igual a la varianza de proceso, para redefinir la posición actual de la partícula de la siguiente forma:

$$x(t + 1) = x(t) + w(t)$$

Aquí, $x(\cdot)$ representa la posición de la partícula en coordenadas cartesianas, pero por simplicidad para la atenuación de la magnitud, se transforma a coordenadas polares, tomando como referencia la posición anterior de la partícula. Se tiene entonces una representación del movimiento de transición de $x(t)$ a $x(t + 1)$ caracterizado en coordenadas polares por (ρ^*, θ^*) , por lo que la atenuación en este caso es de la siguiente forma:

$$\rho_{pred} = \lambda^{t-1} \rho^*$$

Con este ajuste, es posible volver a definir una posición en el plano cartesiano, considerando la magnitud atenuada y la orientación calculada (ρ_{pred}, θ^*) .

Por su parte, la varianza se ajusta de la siguiente forma:

$$\sigma_{pred}^2 = \frac{\sigma_{proceso}^2}{\lambda^{t-1}}$$

Finalmente, se aproxima una distribución de densidad de probabilidad para el riesgo asociado a una partícula y para cada uno de los pasos hasta el horizonte de predicción.

En la Figura 22 se aprecia gráficamente el cambio en la posición de una partícula que es afectada por el Método 2. En A, se aprecia el desplazamiento en coordenadas cartesianas, y en B su transformación a coordenadas polares. Mediante la atenuación a la magnitud del movimiento apreciada en C se caracteriza el valor final de la posición, que en D nuevamente se vuelve al plano cartesiano.

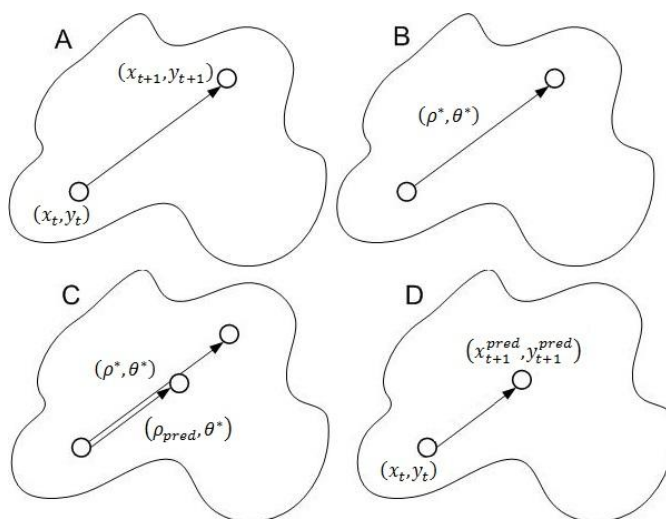


Figura 22: Ejemplo de la caracterización del movimiento de una partícula afectada por el ruido mayoritariamente. La representación del movimiento futuro en el plano cartesiano (A) puede fácilmente transformarse al plano polar (B), donde conservando el valor de la orientación se atenúa la magnitud del movimiento (C), y la posición final atenuada se transforma nuevamente al plano cartesiano (D).

3.3.2.6 Generación de modelo mediante GMM

Finalizando la etapa anterior, para cada partícula se aproxima una distribución, que dado los procedimientos descritos, está condicionada tanto a la última observación disponible como a los movimientos previos de la partícula y la varianza del proceso. Esta distribución representa la zona de riesgo para cada partícula, y cada una de ellas es una representante de la distribución a posteriori del riesgo total del área de interés. Así, la ubicación en el plano de las partículas representará una zona de riesgo que está determinada por la presencia de esta, pues idealmente no debiesen existir partículas en zonas en las que no se han producido delitos.

Como se explica en la Sección 2.6, una distribución a priori puede ser aproximada por un conjunto de muestras de la misma, y similarmente, un conjunto de muestras puede utilizarse para aproximar una distribución a posteriori. En este Trabajo de Título se utiliza un modelo por

mezcla de gaussianas, al igual que en la generación del modelo inicial, pero esta vez centrando cada una de las distribuciones que conforman el modelo en la posición final predicha por la metodología propuesta.

Dado que para cada partícula se tendrá una distribución de riesgo a posteriori, en esta etapa solo basta con generar el modelo por mezcla de gaussianas, esto es, calcular:

$$PDF_{a\ posteriori}(\vec{x}) = \sum_{i=1}^M w_i f_i(\vec{x})$$

Donde $PDF_{a\ posteriori}(\vec{x})$ corresponde a la distribución espacial del riesgo final a posteriori definida en toda el área de interés; $f_i(\vec{x})$ corresponde a la distribución de riesgo a posteriori para la partícula i , y los pesos respectivos cumplen:

$$w_i = \frac{1}{M} \Rightarrow \sum_{i=1}^M w_i = 1$$

Al finalizar esta etapa, se tiene entonces una predicción de la distribución a posteriori del riesgo en base a patrones de movimiento previamente efectuado por cada una de las partículas, condicional al historial de desplazamientos y a la última observación adquirida.

CAPÍTULO 4. CASO DE ESTUDIO Y DISCUSIÓN DE RESULTADOS OBTENIDOS

Este Trabajo de Título se enmarca dentro un gran objetivo: la predicción de focos de criminalidad. Para esto se requiere de un modelo de carácter estocástico y dada la carencia de un modelo, la generación del mismo pasa a ser un objetivo más de este trabajo. En el capítulo anterior se describe una metodología para la generación automática de este modelo y para caracterizar la evolución temporal del mismo, permitiendo la predicción de los focos. Esta propuesta se desarrolla tomando como base el caso de estudio que se presenta en este capítulo.

Para la implementación se utiliza información correspondiente a la comuna de La Florida, Región Metropolitana de Santiago de Chile, entregada bajo acuerdo de confidencialidad por el Centro de Análisis y Modelamiento en Seguridad (*Center for Analysis and Modeling of Security, CEAMOS*) y que fue proporcionada por Carabineros de Chile. En las secciones siguientes se entrega una descripción general de la base de datos que es utilizada para evaluar la metodología propuesta; posteriormente se presentan los resultados obtenidos referentes a la metodología de generación de modelos espaciales de riesgo; finalmente se entregan los resultados referentes a la etapa de caracterización de la evolución temporal del modelo y con ello las predicciones correspondientes, en ambos casos utilizando el método de evaluación de modelos descrito en la Sección 2.10.

4.1 DESCRIPCIÓN DE LA COMUNA ESTUDIADA Y DE LAS BASES DE DATOS UTILIZADAS

Como se mencionó anteriormente, el caso de estudio se enmarca dentro de la comuna de La Florida. La Florida es una comuna de la Región Metropolitana de Santiago de Chile. Forma parte de la Provincia de Santiago y ubica en el sector suroriente de la misma. Limita al sur con la comuna de Puente Alto, al norte con las comunas de Macul y Peñalolén y al oeste con las comunas de La Granja y La Pintana, apreciado en la Figura 23, según los límites comunales señalados en el DFL N° 1-3260 de 1981, publicado en el Diario Oficial del viernes 11 de diciembre de 1981.

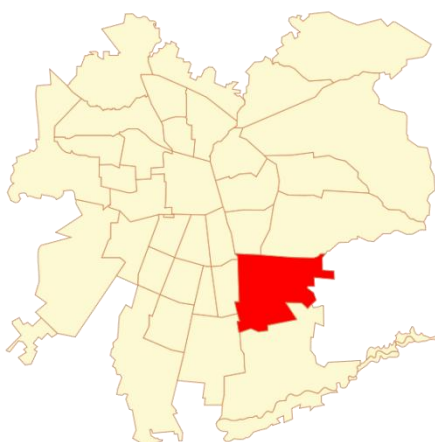


Figura 23: Comuna de La Florida, presentando su ubicación dentro de la ciudad de Santiago de Chile.

Durante los años 1980 y parte de los años 1990 fue la comuna más poblada de Chile, y en la actualidad está habitada principalmente por familias de nivel socioeconómico medio, siendo considerada una de las comunas de la clase media santiaguina. Tiene una superficie de 7.020 hectáreas, es decir 70.2 Km², que corresponde al 3,42 % de la superficie total de Santiago. Además, según los Reportes Estadísticos Distritales y Comunales 2013 [31], en el Censo del año 2002 existía una población total de 365.674 habitantes, con una proyección al 2012 de 397.497 habitantes, de los cuales un 91,3% se consideran como no pobres, con un ingreso promedio comunal de \$951.130 pesos chilenos.

Según información de la Subsecretaría de Prevención del Delito del Ministerio de Interior y Seguridad Pública referente a los Delitos de Mayor Connotación Social (DMCS) [4], en esta comuna existe una tasa de casos policiales de 3.862,4 y una tasa de denuncia de 2.673,2 casos, en ambos casos la tasa se define en casos por cada 100.000 habitantes.

Gracias a la acción de las policías en cada uno de sus roles, existe una gran cantidad de información que es recolectada en el día a día por Carabineros de Chile, tanto en base a las denuncias presentadas por la ciudadanía en la policía como por los delitos con aprehendidos en flagrancia. Ellos disponen de una extensa base de datos, una pequeña parte de la cual fue proporcionada para la realización de este Trabajo de Título, en particular aquella referente a la comuna de La Florida.

Esta información consiste en:

- Información georeferenciada de servicios ubicados en la comuna: base de datos con 4262 registros a la fecha 22 de Agosto del 2012. Estos servicios corresponden a cualquier punto característico de cualquier asentamiento urbano, como por ejemplo: almacenes, bancos, bares, estaciones de bomberos, botillerías, cajeros automáticos, comisarías, centros comerciales, colegios, parques, hospitales, consultorios, solo por dar algunos ejemplos. La información utilizada corresponde al tipo de servicio y las coordenadas en sistema GPS (latitud y longitud).
- Información georeferenciada de eventos delictuales ocurridos en la comuna: base de datos correspondiente al período de tiempo comprendido entre los meses de septiembre del 2012 a febrero del 2013. Con un total de 23109 registros, siendo los más frecuentes aquellos relacionados con infracciones a la ley del tránsito, seguido por hurto, infracciones a la ley de alcoholes y robo con fuerza. Se decide modelar este último tipo de delito (robo con fuerza), del cual se tiene un total de 2240 eventos que deben ordenarse cronológicamente. Se divide entonces en un conjunto de entrenamiento con los primeros 1870 eventos, uno de prueba con los siguientes 185 eventos y uno de validación con los siguientes 185 (aproximadamente el 10% de los datos totales en cada conjunto de prueba y validación). La información utilizada corresponde a la fecha y hora del evento, el tipo de delito, y las coordenadas en sistema GPS (latitud y longitud).

Actualmente existen variados programas computacionales de Sistemas de Información Geográfica (GIS) que facilitan el trabajo de visualización [32], pero por motivos presupuestarios no son utilizados en este Trabajo de Título. A modo de compensación, para asistir en la visualización y comprensión de los resultados, se utiliza un plano de la comuna que puede obtenerse gratuitamente desde el servicio web openstreetmap [33] o bien de cualquier software GIS, como se aprecia en la Figura 24.

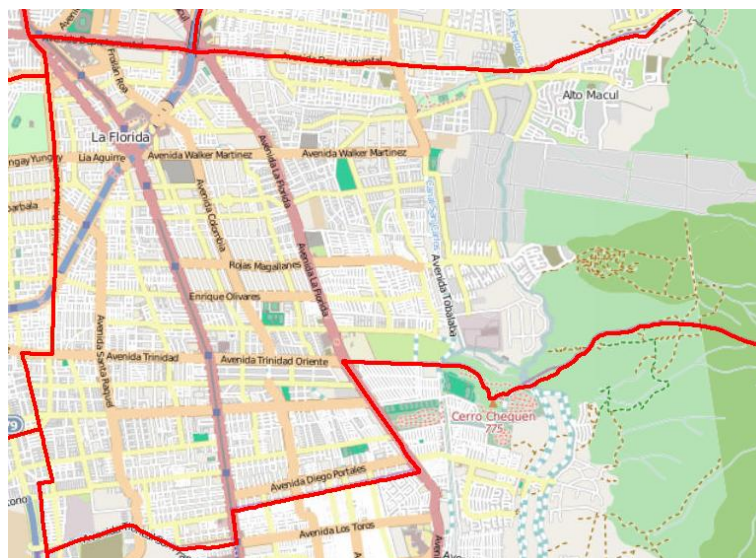


Figura 24: Límites comunales de La Florida y sus avenidas principales.

4.2 RESULTADOS REFERENTES A LA METODOLOGÍA DE GENERACIÓN AUTOMÁTICA DE MODELOS ESPACIALES DE RIESGO

Para la creación de modelos espaciales de riesgo es necesario definir el área de interés y con ella asociar incidentes delictuales y servicios (Sección 3.1). En la Figura 25 se muestran tanto los incidentes (izquierda) como los servicios (derecha) ubicados en el área de interés, presentados separadamente para mejor comprensión de su distribución espacial.

Salvo por la ausencia de servicios e incidentes en la zona rural de la comuna (apreciada en cada gráfico hacia la derecha, aproximadamente entre las longitudes $-70,54^\circ$ y $-70,50^\circ$), es difícil encontrar algún patrón entre ellos, por lo que se puede asegurar que la generación de un modelo probabilístico de la distribución de riesgo no es tan sencillo de realizar mediante la teoría de Hot Spot (Sección 2.2), al menos no a gran escala como en este caso de estudio.

Además, en esta figura se presentan solo los delitos tipificados como robo con fuerza, y se consideran todos los servicios de los que se tiene registro hasta la fecha (ver Sección 4.1 para mayor información al respecto).

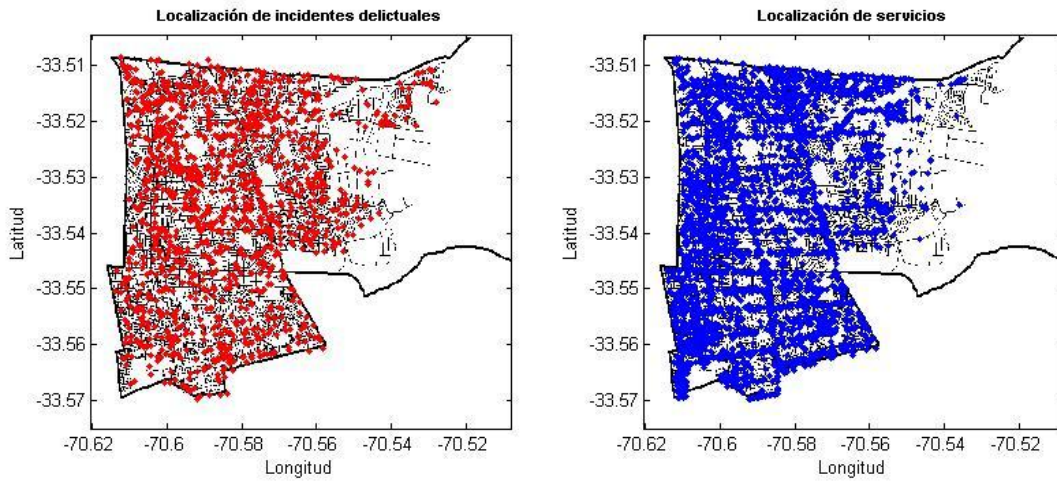


Figura 25: Localización de incidentes delictuales y de servicios en el área de interés. A simple vista no es posible identificar focos de criminalidad, así como tampoco se aprecia una clara correlación entre ambos.

Esta información es utilizada para la generación de los modelos probabilísticos según la metodología propuesta en la Sección 3.2 y cuyo desempeño depende principalmente de dos parámetros de diseño: la cantidad de servicios utilizados (más específicamente su distribución en el plano) y el rango de eventos utilizados para la aproximación de la función de densidad de probabilidad.

En la Figura 26 se presentan tres modelos correspondientes al uso de 222 servicios del tipo ‘Botillería’, utilizando 3, 5 y 7 cuadras de rango para su creación. En la Figura 27 se presenta un caso diferente, presentando tres modelos para los servicios etiquetados como ‘Gran Tienda Comercial’, con un total de 11 elementos, y en los que se utilizan 3, 5 y 7 cuadras como rango.

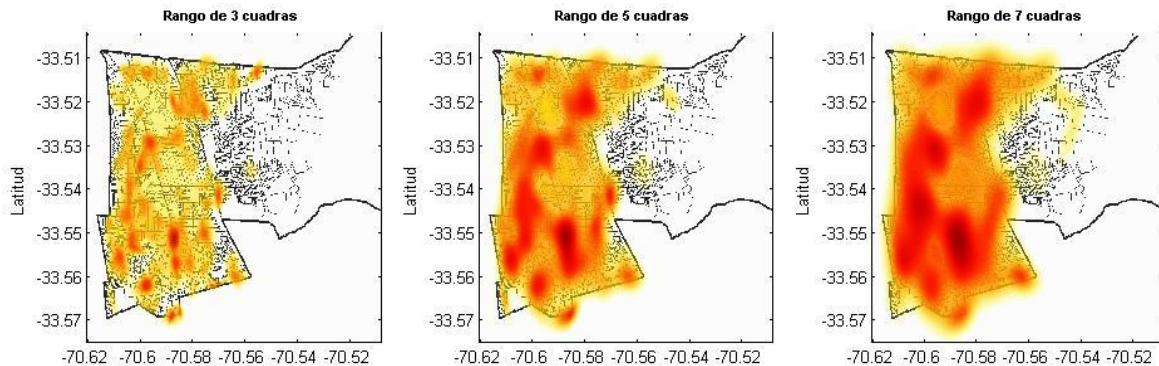


Figura 26: Modelos generados para servicios tipo ‘Botillería’, considerando 3, 5 y 7 cuadras respectivamente para su creación.

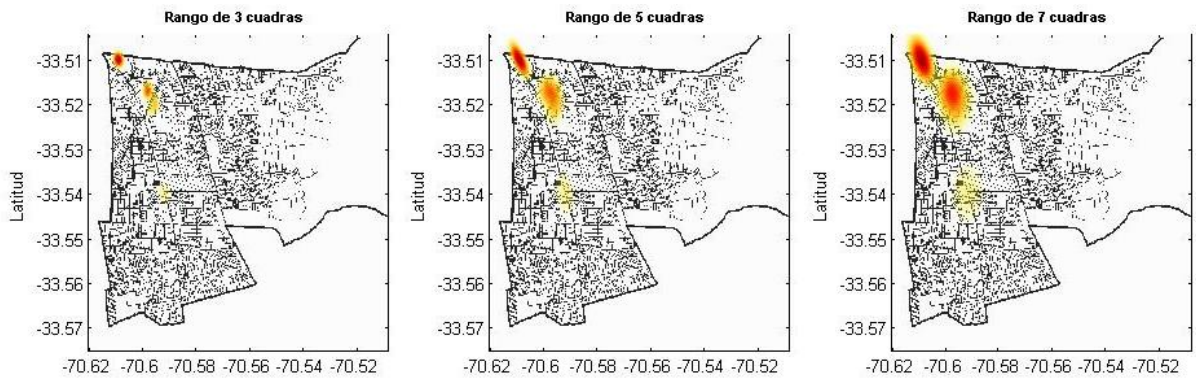


Figura 27: Modelos generados para servicios del tipo 'Gran Tienda Comercial', considerando 3, 5 y 7 cuadras respectivamente para su creación.

En ambas figuras es posible apreciar que el aumento en el rango utilizado entrega mayor incertidumbre que no necesariamente es una mejora en el modelo; por otro lado, un rango menor no es representativo de toda el área de interés. Si se consideran rangos inferiores a las 3 cuadras (considerando el largo de una cuadra promedio como 125 metros [34]) el problema radica en que no existe suficiente información para poder generar una estadística en torno al servicio (entiéndase, calcular la varianza de los datos asociados a ese servicio), por lo que rangos menores no se consideran. Los casos en los que se utilizan rangos superiores a siete cuadras, tampoco se incluye en el análisis puesto que estos modelos sobreestiman el verdadero riesgo asociando demasiados eventos a un solo servicio, más aún considerando la lejanía en el plano, que finalmente solo agrega una gran incertidumbre en el modelo.

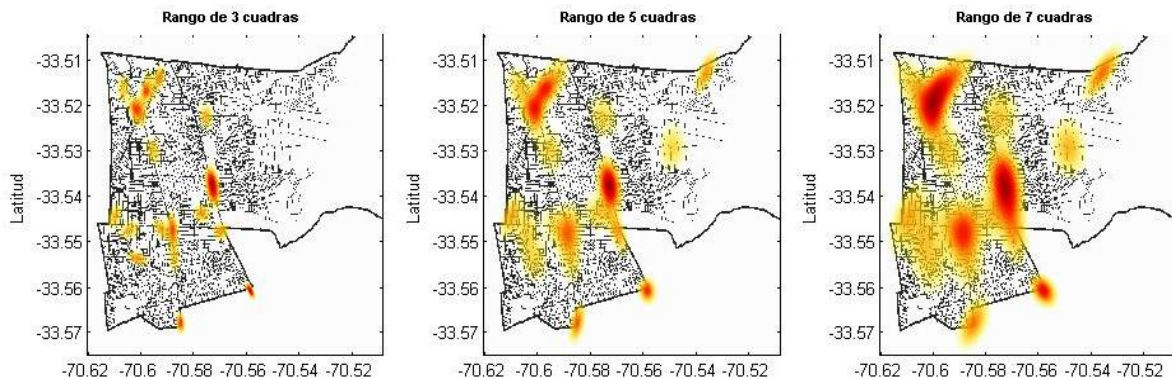


Figura 28: Modelos generados para servicios del tipo 'Bar y Pub', considerando 3, 5 y 7 cuadras respectivamente para su creación.

Ahora bien, en la Figura 28 se muestran tres modelos para los servicios etiquetados como ‘Bar y Pub’, con un total de 27 elementos, y en los que se utilizan 3, 5 y 7 cuadras como rango.

Gracias a los modelos presentados en la Figura 27 y Figura 28 es posible apreciar la relación que debe existir entre el área de interés y la cantidad de servicios utilizados y su localización. En ambos casos se tiene una pequeña cantidad de servicios para crear el modelo, pero su dispersión en el plano es diferente. En el caso de las grandes tiendas comerciales estas se encuentran poco dispersas, por lo que el modelo no cubre gran parte del área de interés; en el caso de los bares y pubs, que se encuentran un poco más distribuidos en el plano, representan un mayor porcentaje del área de interés.

La cantidad de servicios que se consideran para la creación del modelo influye directamente en el tiempo de procesamiento requerido para crearlo. En la Figura 29 se puede apreciar que existe una relación casi lineal (al menos para los 4000 primeros datos procesados) que relaciona el tiempo de procesamiento para crear el modelo con la cantidad de servicios que se utilizan.

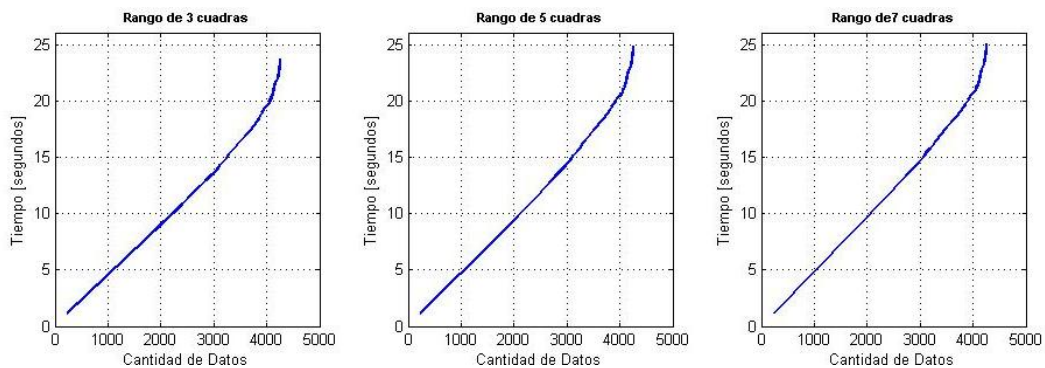


Figura 29: Gráficos comparativos de los tiempos de procesamiento para la generación del modelo, considerando un rango de 3, 5 y 7 cuadras respectivamente. No se aprecia una diferencia sustancial en ellos.

Para procesar el total de 4.262 servicios se tienen tiempos de procesamiento de 23,75 [s] para un modelo de 3 cuadras de rango, 24,94 [s] para el modelo de 5 cuadras y 25,12 [s] para el modelo de 7 cuadras. Dado que esta etapa se lleva a cabo solo una vez por la metodología propuesta, se decide crear el modelo a priori de la distribución del riesgo con la totalidad de servicios conocidos.

Ahora, el criterio del tiempo de procesamiento solo entrega una guía de cuántos servicios utilizar, pero no presenta un criterio claro para decidir qué rango utilizar. Para esto se utiliza entonces el método de evaluación definido en la Sección 2.10, utilizando el conjunto de prueba descrito en la Sección 4.1, cuidando el orden cronológico. Es decir, el uso de este conjunto de prueba no altera el orden de ocurrencia de los eventos, y como se verá en la sección siguiente, puede utilizarse además para pruebas de la evolución temporal.

Se utiliza entonces este conjunto para evaluar el modelo, variando tanto el rango de cuadras utilizadas para crear el modelo (en cada fila), como el tamaño de la grilla utilizada para evaluar (en cada columna). Se utilizan grillas cuadradas para definir cada una de las áreas en las que se evalúa el modelo, con la consideración pertinente de que el área de interés no necesariamente puede dividirse exactamente en un tamaño fijo, por lo que algunas grillas en la frontera podrían ser más pequeñas.

En la Tabla 2 se presentan los resultados de la evaluación del modelo, indicándose para cada combinación el valor del indicador AUC para cada modelo generado.

Tabla 2: AUC para los modelos generados, variando el rango de modelación de 3 a 7 cuadras, y el tamaño de la grilla de 1 a 10 cuadras.

	$g = 1$	$g = 2$	$g = 3$	$g = 4$	$g = 5$	$g = 6$	$g = 7$	$g = 8$	$g = 9$	$g = 10$
$r = 3$	0,87	0,87	0,87	0,87	0,86	0,87	0,87	0,88	0,86	0,85
$r = 4$	0,87	0,87	0,88	0,87	0,87	0,87	0,87	0,88	0,86	0,85
$r = 5$	0,87	0,87	0,87	0,87	0,87	0,86	0,87	0,88	0,86	0,85
$r = 6$	0,87	0,87	0,87	0,87	0,87	0,87	0,87	0,88	0,86	0,85
$r = 7$	0,87	0,87	0,87	0,87	0,87	0,87	0,87	0,87	0,86	0,85

Recordando que mientras mayor sea el valor de este indicador mejor será el modelo generado, en la Tabla 2 se encuentran destacados los valores máximos para cada modelo. Se aprecia que los valores máximos se alcanzan para una grilla de tamaño 8 cuadras para todos los modelos, con la excepción del modelo generado con un rango de 4 cuadras en que también se alcanza un valor máximo con una grilla de tamaño 3 cuadras. Sin embargo, se considera que este valor es una excepción a la regla, pudiendo deberse netamente al azar y no a un mejor esquema de evaluación.

El criterio para decidir el rango de la creación el modelo se basa entonces en la maximización del AUC, pero debe considerarse también un aspecto práctico que puede ocurrir en otras realizaciones. Si bien el AUC podría ser similar para cada modelo, se debe siempre considerar la grilla de menor tamaño por sobre aquella grilla de mayor tamaño. Por ejemplo, si para un modelo se maximiza el AUC utilizando una grilla de tamaño 10 cuadras, y para otro modelo se maximiza considerando una grilla de tamaño 8, este último esquema de evaluación es el que se debe considerar. En el papel puede no parecer una gran diferencia, pero en la práctica esto se traduce en un área de monitoreo mayor, y los patrullajes pueden no resultar tan efectivos como si lo serían en áreas menores.

En este caso de estudio la situación anterior no se evidencia, y se opta por utilizar como criterio la maximización del AUC que se obtiene considerando una grilla de tamaño 8 cuadras, y un modelo generado a partir de un rango de 6 cuadras, para efectos de representar un mayor porcentaje del área de interés. En la Figura 30 se aprecia el modelo probabilístico respectivo, con $AUC=0,8754$ y que en adelante se utiliza para los análisis y resultados futuros.

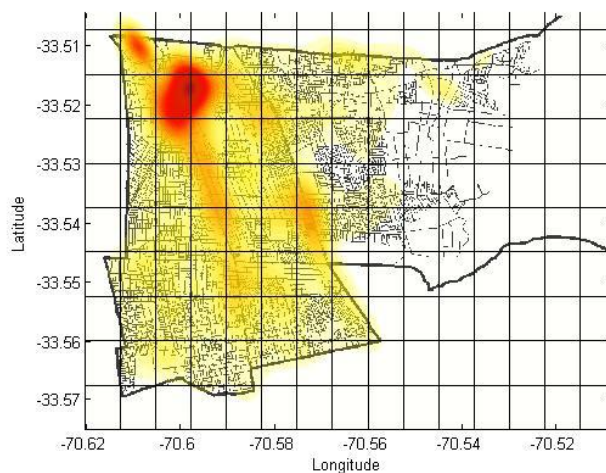


Figura 30: Distribución espacial del riesgo, considerando los valores que optimizan criterio de AUC y distribución de recursos: rango 6 cuadras y grilla de 8 cuadras.

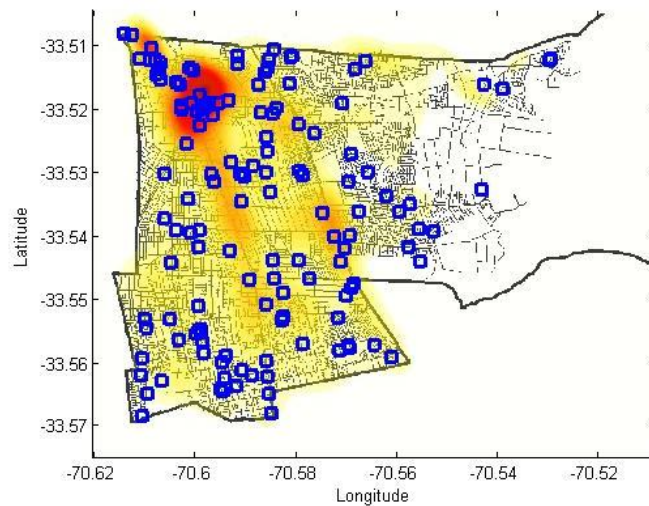


Figura 31: Esquema representativo del modelo de riesgo generado (superficie en escala de colores de blanco a amarillo y rojo) y la distribución espacial de los eventos del conjunto de prueba (cuadrados azules). Se aprecia que la mayoría de estos eventos se presentan en las zonas catalogadas como "de alto riesgo" o cerca las mismas.

En la Figura 31 se aprecia una comparación gráfica entre la distribución de riesgo generada mediante la metodología descrita anteriormente y los eventos de un conjunto de prueba. Es posible apreciar que muchos de estos eventos ocurren en aquellas zonas de alto riesgo (zonas de color rojo y amarillo intenso) o cerca de éstas, mientras que son bastante escasos en las zonas de bajo riesgo (blanco y amarillo claro). Esto indica que el modelo caracteriza de buena manera los eventos futuros, mostrándose entonces cierta relación entre ellos y los datos pasados. Ahora, si bien hay zonas catalogadas como de bajo riesgo pero que aún así presentan eventos delictuales (ver cluster en latitud $-33,54^\circ$ y longitud $-70,55$), se espera que las iteraciones del algoritmo de evolución temporal permitan caracterizar este foco que la modelación previa no caracterizó.

4.3 RESULTADOS REFERENTES A LA METODOLOGÍA DE CARACTERIZACIÓN DE LA EVOLUCIÓN TEMPORAL DE MODELOS ESPACIALES DE RIESGO

Tal como se describe en la Sección 3.2.2 y representado en la Figura 15 y la Figura 17, la metodología de caracterización de la evolución temporal de los modelos de riesgo puede dividirse en dos etapas, offline y online respectivamente, haciendo uso del modelo probabilístico obtenido anteriormente y presentado en la Figura 30.

El primer paso para la caracterización de la evolución temporal comienza por la obtención de muestras de la distribución de densidad de probabilidad que sean representativas de la misma: el muestreo de importancia. Como se explica en la Sección 2.6, dada la inviabilidad de tomar muestras de la distribución a priori, es necesario definir otra distribución de la que sea más sencillo el muestreo. Si bien puede definirse sin mayores problemas una distribución uniforme en toda el área de interés y muestrear a partir de ella, se propone realizar el muestreo de importancia a partir de una grilla y remuestreo, que entrega resultados equivalentes en la teoría y mejores en la práctica.

El primer inconveniente que se presenta a la hora de muestrear a partir de una distribución uniforme bivariada es la cantidad de partículas (muestras) a utilizar. En la Figura 32 se aprecia que el muestreo de importancia al considerar un número muy bajo de partículas no resulta representativo de la distribución inicial, posicionando las partículas lejos de los focos modelados (ver Figura 30). Esta situación se produce principalmente por la aleatoriedad del proceso, que al tener pocas realizaciones no garantiza representar completamente el área de interés. Esta situación se revierte al considerar un número mayor de partículas, como se muestra en la Figura 33 y Figura 34; sin embargo en esta última, que considera un total de 500 partículas, se comienza a sobreestimar el riesgo (dado por la presencia de partículas en zonas poco riesgosas) y por ende el tiempo de procesamiento de la etapa siguiente (online) será mayor.

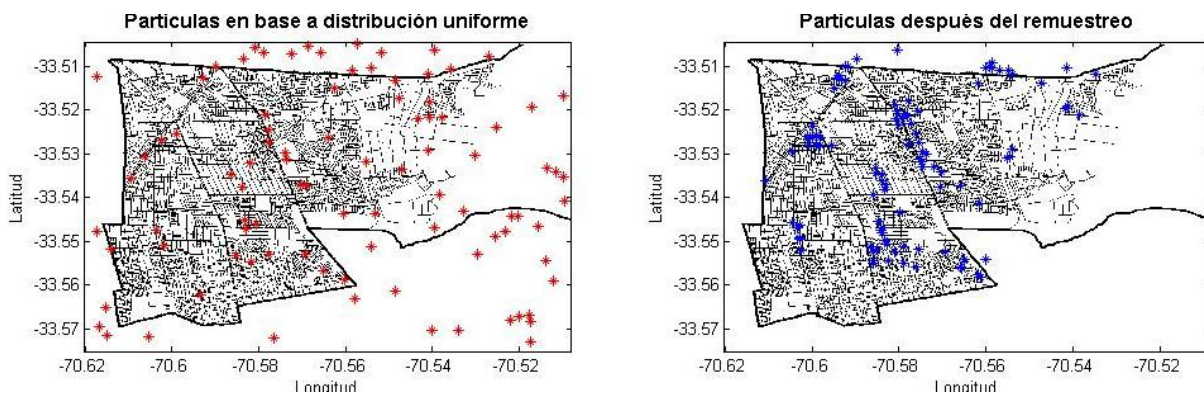


Figura 32: Comparación entre posiciones de 100 partículas en base a una distribución uniforme y posterior al remuestreo.

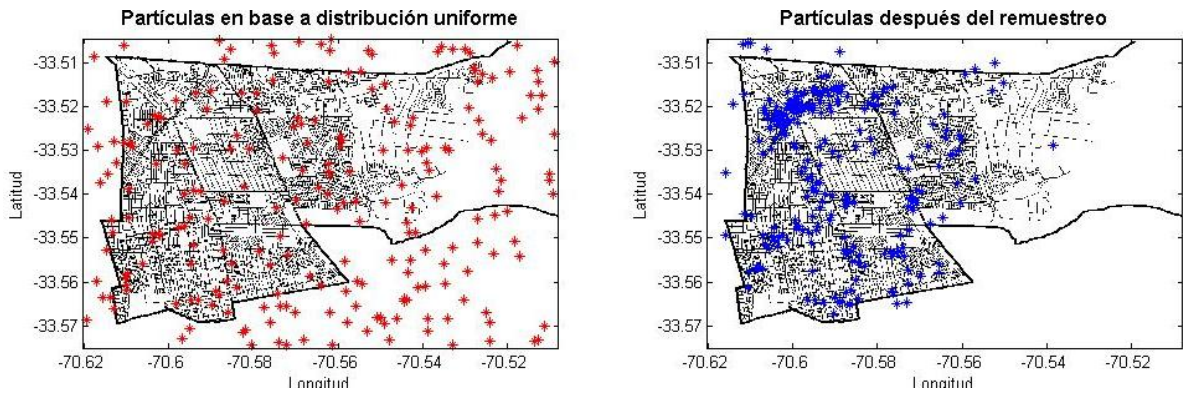


Figura 33: Comparación entre posiciones de 250 partículas en base a una distribución uniforme y posterior al remuestreo

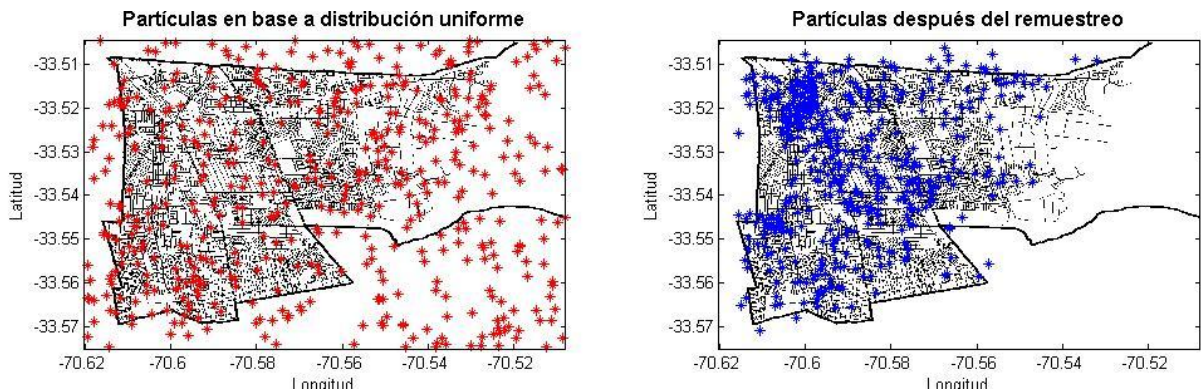


Figura 34: Comparación entre posiciones de 500 partículas en base a una distribución uniforme y posterior al remuestreo

El segundo inconveniente se relaciona con la confiabilidad de las muestras generadas: si se considera un número medianamente alto, a modo de evitar la falta de representantes en gran parte del área de interés, pero al mismo tiempo no sobrecargue el tiempo de procesamiento, no se puede garantizar mediante este método que las partículas aproximarán la distribución a priori. Considérese por ejemplo los casos de la Figura 35 y Figura 36. Si bien ambos casos presentados se tiene un total de 370 partículas, la generación aleatoria resulta en dos distribuciones muy diferentes, omitiéndose en el primer caso un foco en la zona norponiente, y en el segundo caso se obtiene una escasa representatividad en la zona central de la comuna (en ambos casos, marcados con un círculo rojo).

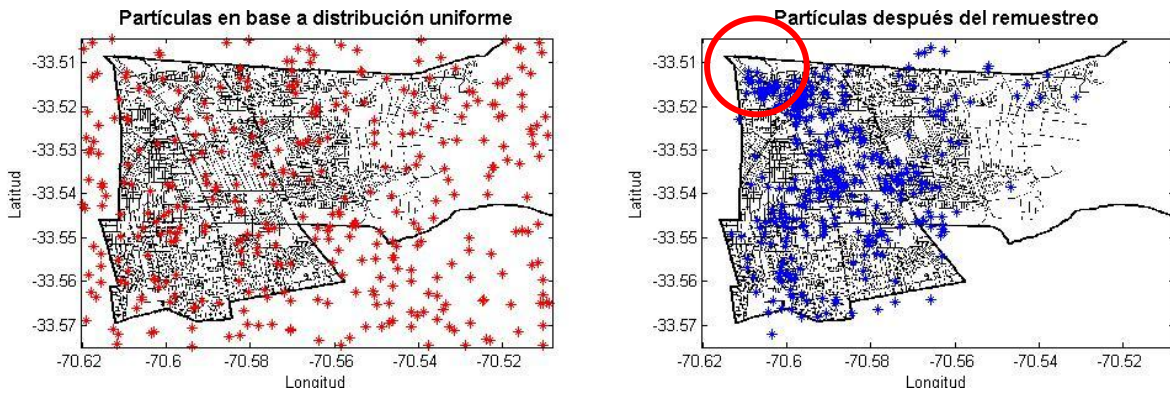


Figura 35: Posición de partículas, donde se puede apreciar que no existe representatividad en el foco de zona norponiente.

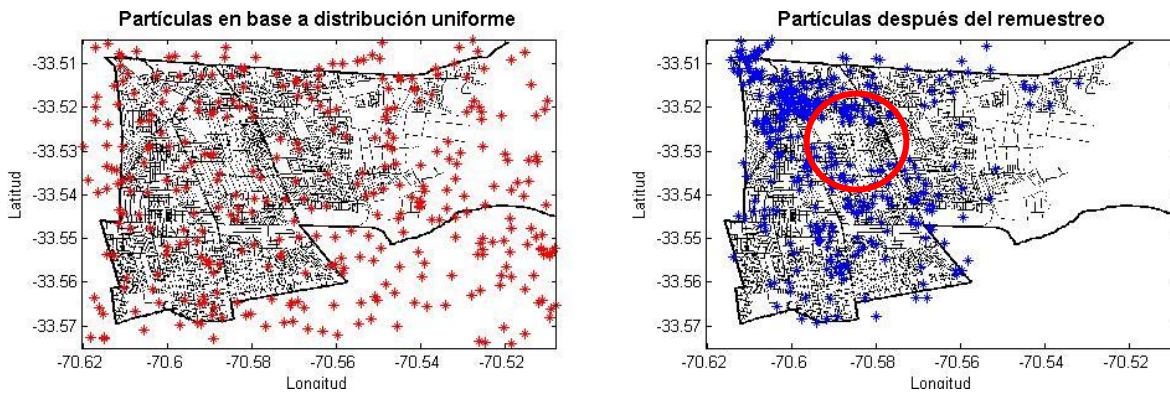


Figura 36: Posición de partículas, donde se puede apreciar que existe escasa representatividad en la zona central.

Revisando la propuesta de muestrear a partir de una grilla y utilizar el algoritmo de remuestro presentado en la Sección 2.8, es posible evitar los dos inconvenientes descritos. En primer lugar, el número de partículas queda determinado por el tamaño de la grilla, que resulta más intuitivo por el hecho de que entrega una noción del área que “explora” cada una de las partículas (la diferencia entre una grilla con separación de una cuadra entregará un mayor número de partículas y por ende mayor representatividad de toda el área de interés que una grilla con separación de 10 cuadras).

En segundo lugar, al tener una inicialización fija determinada por la grilla, la aleatoriedad del proceso depende exclusivamente del proceso de remuestreo, que utiliza información del peso inicial de la partícula dado por la distribución a priori para re posicionar las partículas en el plano, a diferencia del esquema anterior que depende también de las muestras de la distribución uniforme.

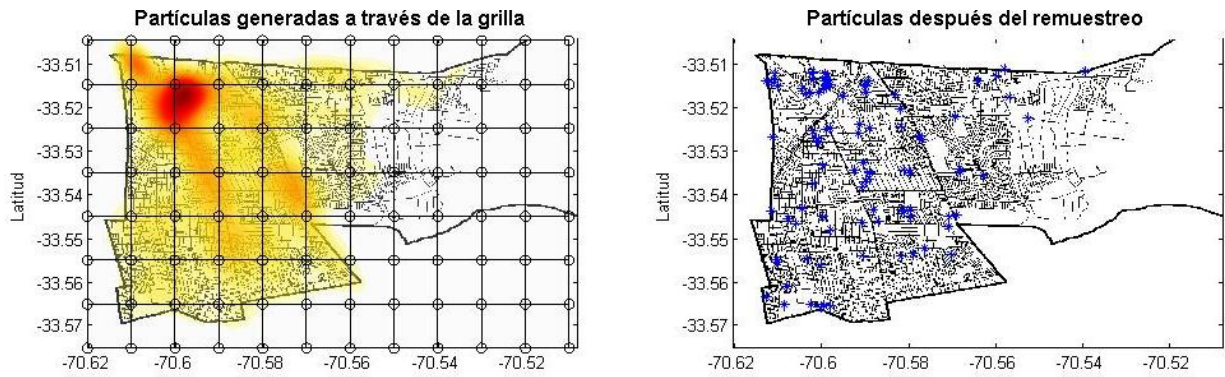


Figura 37: Comparación entre posiciones de 96 partículas en base a grilla de 10 cuadras y posterior al remuestreo.

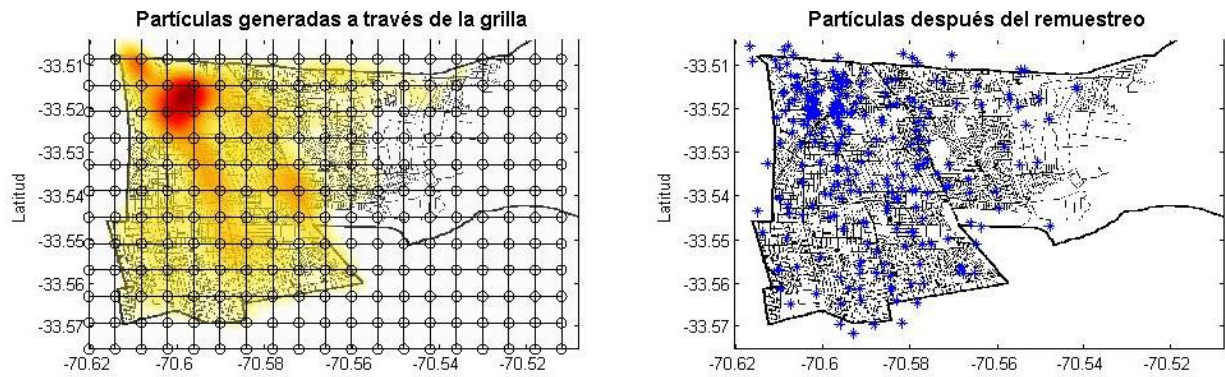


Figura 38: Comparación entre posiciones de 247 partículas en base a grilla de 6 cuadras y posterior al remuestreo.

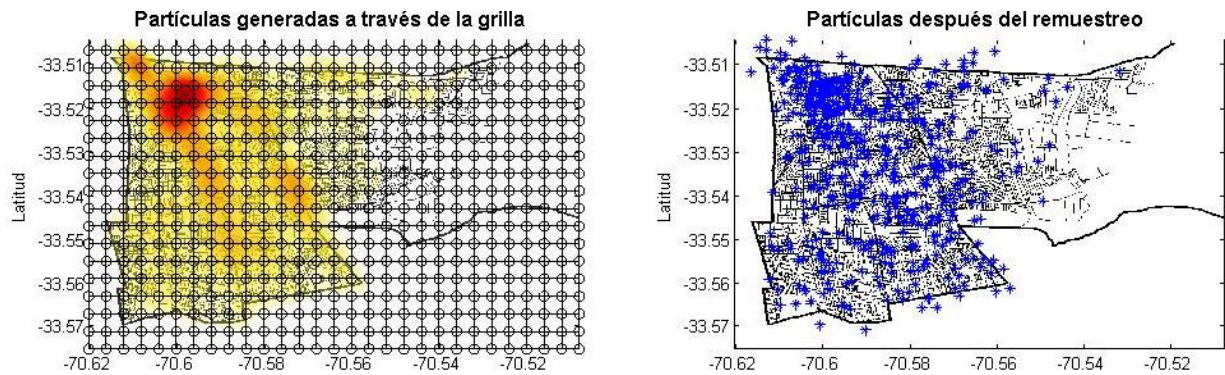


Figura 39: Comparación entre posiciones de 551 partículas en base a grilla de 4 cuadras y posterior al remuestreo.

Si se compara la Figura 32 con la Figura 37, se puede apreciar que en el caso de la grilla el principal foco de la comuna sigue teniendo representatividad, a diferencia del muestreo a partir de una distribución uniforme, en que el foco se pierde; además, si se compara la Figura 34 con la

Figura 39 se puede ver que si bien existe una sobreestimación del riesgo, la mayor parte de las partículas se concentra en los focos, lo que no sucede en el esquema anterior.

Como se puede ver, utilizar una grilla produce mejores resultados, principalmente por la menor aleatoriedad del proceso de selección de partículas que permite una mejor representación de toda el área de interés, y ayudando a definir con más información el número de partículas. Por estas razones se utiliza este método para el muestreo de importancia, considerando una grilla de 5 cuadras que produce un total de 368 partículas (suficiente para caracterizar gran parte del área de interés sin sobrecargar el proceso online, como se aprecia en la Figura 40).

Las posiciones de las partículas serán utilizadas para aproximar una nueva distribución espacial del riesgo, en la que las zonas de alta probabilidad estarán dadas por la acumulación de partículas en esa zona específica en lugar del peso de la partícula. En la Figura 41 se aprecian dos distribuciones espaciales del riesgo; en la distribución de la izquierda se muestra la distribución a priori calculada anteriormente (Sección 4.2), y en la distribución de la derecha se muestra la aproximación obtenida mediante el muestreo de importancia y el uso de un modelo por mezcla de gaussianas (Sección 3.3.2). Si bien ambas se aprecian diferentes, evaluando ambos modelos se obtiene en el primer caso un $AUC=0,8754$ y en el segundo caso $AUC=0,8646$, por lo que ambos son bastantes similares considerando el uso que se les dará y los resultados buscados con el algoritmo.

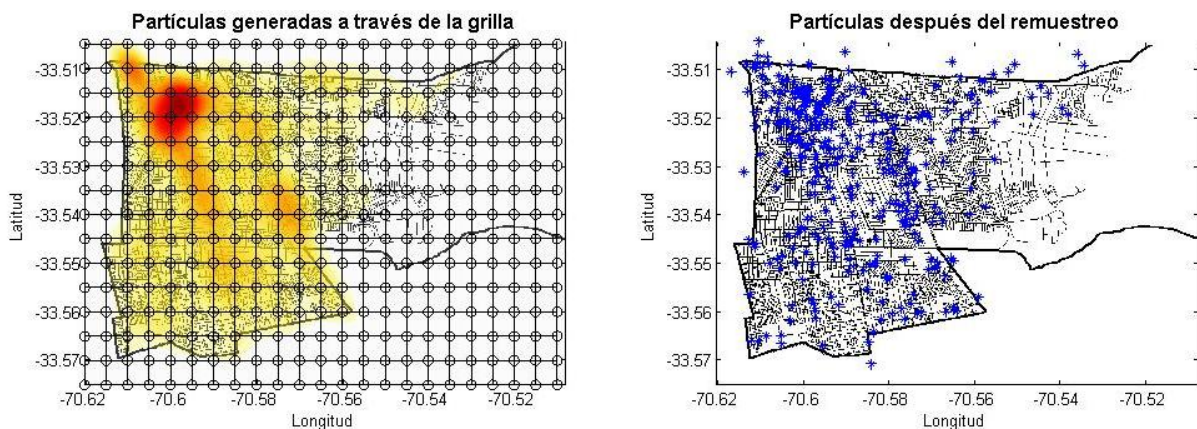


Figura 40: Comparación entre posiciones de 368 partículas en base a grilla de 5 cuadras y posterior al remuestreo

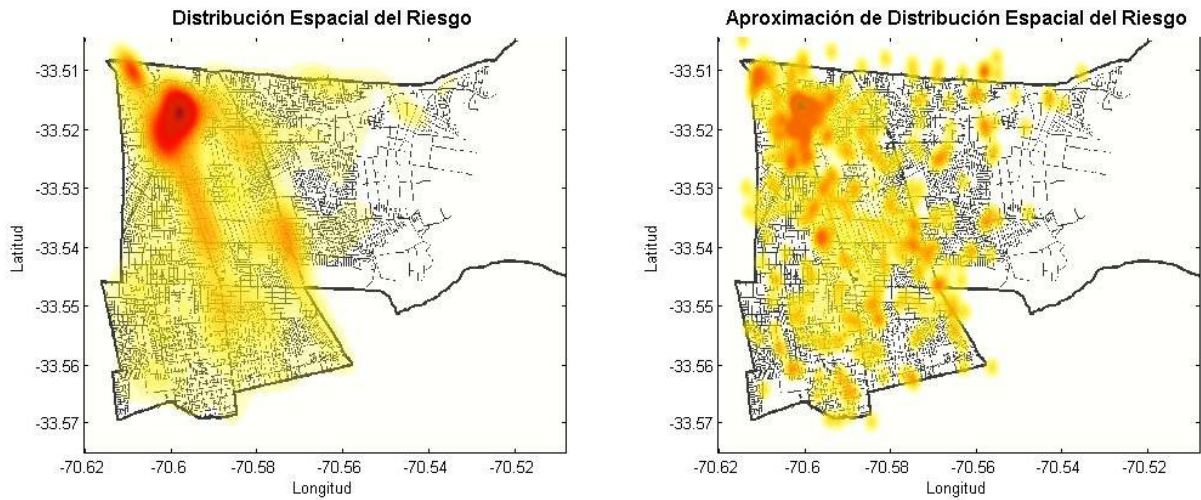


Figura 41: Comparación entre la distribución a priori del riesgo y su aproximación mediante muestreo de importancia (IS) y modelo por mezcla de gaussianas (GMM).

La segunda parte del proceso offline corresponde a la definición del ajuste del modelo dinámico, que corresponde a definir la función $f(d)$, dependiente de la diferencia d entre la observación y una partícula. Como se explica en la Sección 3.3.2, Segundo Bloque Offline, la matriz diagonal Σ es un parámetro de diseño y depende del área de interés, pero principalmente depende de cuánto se desea que una partícula se aproxime a una nueva observación. En otras palabras, define cuánto afecta una nueva observación al entorno de esta.

En la Figura 42 es posible observar el efecto de la matriz de covarianza Σ en el movimiento de 8 partículas de prueba en el gráfico "Posición Inicial". Dichas partículas están marcadas por un círculo rojo, y la observación actual con un triángulo azul.

Cuando se utiliza una varianza muy grande (entiéndase, grande para el problema en cuestión, en este caso se considera una varianza cuatro órdenes de magnitud mayor al ruido de proceso), las partículas que se encuentran lejos de la observación se acercan a ella, lo que se traduce empíricamente hablando en una convergencia instantánea de todas las partículas a la nueva observación, perdiéndose la representatividad y el sentido de la estimación. Por el contrario, si la varianza es muy pequeña (en este caso de la misma magnitud que el ruido de proceso) no se aprecia ningún acercamiento a la observación, incluso para aquellas partículas cercanas es imposible definir una tendencia.



Figura 42: Comparación entre el ajuste producido por diferentes varianzas. En la imagen de la izquierda se aprecian las posiciones iniciales de las partículas; en la imagen central se aprecia el efecto de una matriz de covarianza de valores muy altos, produciendo un acercamiento de todas las partículas independiente de su lejanía a la observación; en la imagen de la derecha se aprecian los efectos de una matriz de covarianza con valores muy pequeños, sin notarse un claro desplazamiento incluso de las partículas cercanas.

En la Figura 43 se muestra el caso en el que se utiliza una varianza un orden de magnitud mayor a la varianza de proceso, pudiendo notarse que las partículas cercanas a la nueva observación se acercan, pero las lejanas se mantienen solo afectadas por el ruido de proceso.

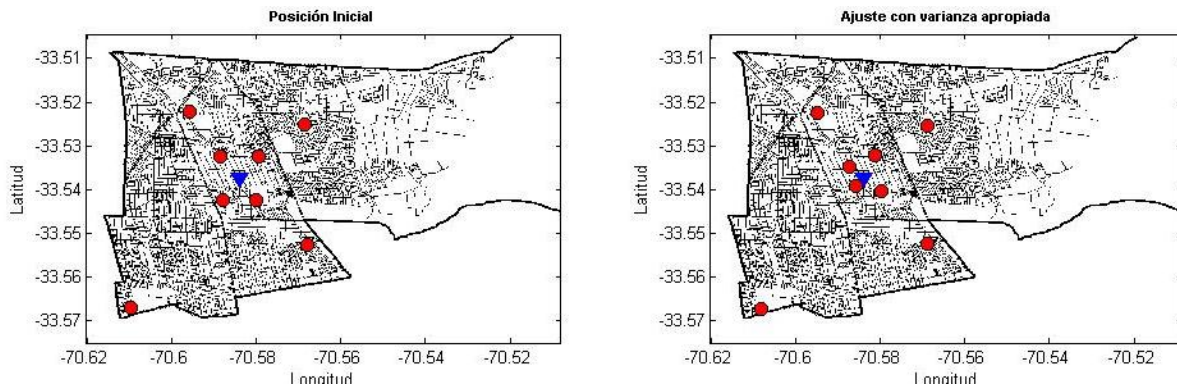


Figura 43: Ajuste en la posición de las partículas producido por una varianza apropiada. En la imagen de la izquierda se muestra la posición inicial de las partículas, mientras que en la imagen de la derecha se aprecian los efectos del uso de una matriz de covarianza apropiada: las partículas lejanas se mantienen lejos, y las cercanas se acercan aún más a la observación.

Dado ajuste presentado, en adelante se considera una matriz de covarianza definida por:

$$\Sigma = \begin{bmatrix} 1,5 \cdot 10^{-5} & 0 \\ 0 & 1,5 \cdot 10^{-5} \end{bmatrix}$$

A partir de esta matriz de covarianza es posible aproximar el umbral de acción u , que indica el límite para el cual el efecto de la observación deja de ser notorio en los desplazamientos de las partículas, que en este caso es igual a 8 cuadras. Es necesario señalar que es mejor sobreestimar este valor (con el consiguiente mayor gasto de recursos computacionales) pero asegurando que todas las partículas efectivamente afectadas por la observación sean procesadas mediante un filtro RLS, mientras que trabajar con un umbral menor puede traducirse en pérdida de información valiosa proporcionada por el movimiento de partículas y las nuevas observaciones incorporadas.

Ambos parámetros (la matriz de covarianza y el umbral de acción) son utilizados en adelante para modificar las posiciones de las partículas según el modelo dinámico descrito en la Sección 3.3.2, Primer Bloque Online.

En la Figura 44 se aprecian dos cambios consecutivos en las posiciones de las partículas (en azul), afectadas por dos nuevos incidentes (en rojo). Para mayor comprensión, la zona de relevancia desde donde y hacia donde se produce el acercamiento de las partículas se ha demarcado con un círculo rojo, permitiendo comparar visualmente con las posiciones iniciales el cambio producido por el modelo dinámico.

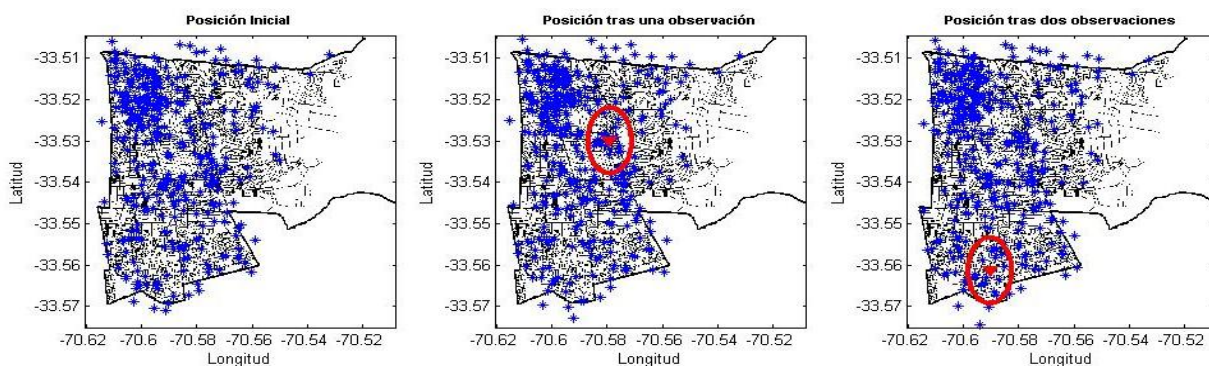


Figura 44: Ejemplo del movimiento de las partículas afectadas por la presencia de una observación. Comparando la imagen de la izquierda con las dos siguientes, se puede apreciar el movimiento y acercamiento de las partículas en las zonas de interés marcadas con un círculo rojo.

No se entregan mayores resultados de esta etapa pues el movimiento es similar para cada iteración. Una vez realizada la actualización de las posiciones de las partículas se aproxima

mediante GMM una distribución espacial del riesgo, en adelante PDF a posteriori (como se explica en la Sección 3.2.2, Primer Bloque Online).

Utilizando nuevamente el conjunto de prueba descrito, y el método de evaluación, es posible calcular el AUC para cada uno de los modelos generados al hacer la evolución artificial de las partículas mediante el modelo dinámico. Para hacer la evaluación se considera el modelo generado hasta la observación τ , y se evalúa con las observaciones restantes, entiéndase, desde $\tau + 1$ hasta 185, teniendo el cuidado de considerar suficientes datos para la evaluación. Considerando entonces 100 iteraciones dadas por los 100 primeros incidentes ocurridos, los resultados son presentados en la Figura 45: aquí, se tiene un AUC promedio de 0,8571 con varianza $1,18 \cdot 10^{-4}$, indicando que los modelos generados predicen de manera bastante certera la ocurrencia de nuevos eventos (recordando, como se explica en la Sección 2.10, que mientras mayor sea el AUC mejor es la predicción del modelo).

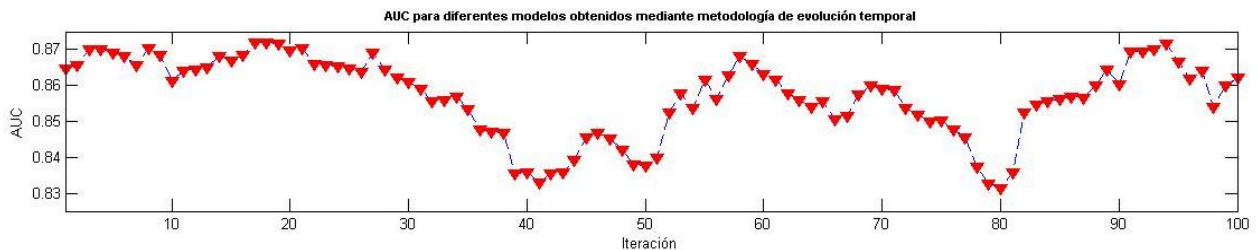


Figura 45: AUC para los modelos obtenidos mediante la metodología de evolución temporal con el modelo dinámico propuesto. Se aprecia que incluso en los momentos en que se alcanzan los valores mínimos, estos siguen siendo superiores a 0.83, demostrando así que el modelo sigue representando de buena forma los incidentes futuros.

Utilizando el mismo conjunto de prueba se presenta en la Figura 46 la distribución de riesgo obtenida tras las 100 realizaciones del algoritmo de evolución temporal, comparándola espacialmente con los siguientes 85 eventos delictivos. Es posible apreciar que aquellas zonas de alto riesgo que presentaron nuevos incidentes en sus cercanías continúan caracterizadas como “de alto riesgo”, mientras que aquellas zonas presentadas en la Figura 31 como “de bajo riesgo”, pero en las que aún así ocurrían incidentes en sus cercanías, ahora tienen algunas partículas que las representan y por ende la probabilidad de ocurrencia de delitos en esa zona aumenta, quedando entonces mejor caracterizadas que antes.

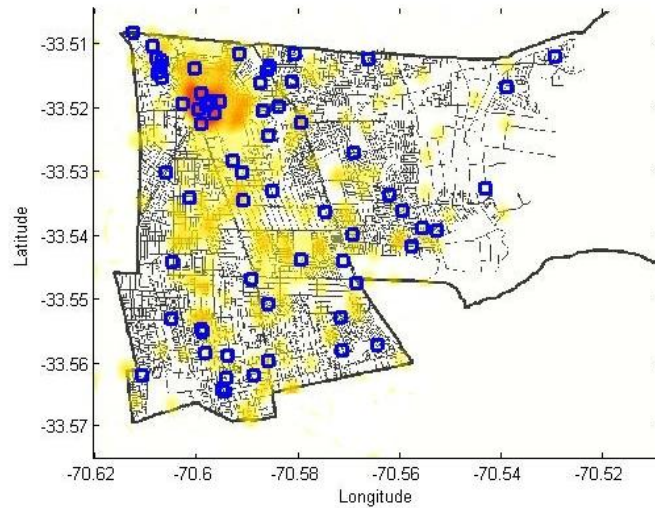


Figura 46: Esquema comparativo entre la distribución espacial de riesgo (amarillo y rojo) posterior al algoritmo de evolución temporal propuesto y eventos futuros del conjunto de prueba (azul). Se aprecia que muchos eventos continúan ocurriendo en las zonas de alto riesgo, pero a diferencia del modelo presentado en la Figura 31 se comienzan a apreciar nuevos focos en las zonas no caracterizadas previamente.

El Segundo Bloque Online utiliza el umbral definido para dividir las partículas en dos conjuntos, siguiendo cada una de ellas caminos diferentes dependiendo de a cuál pertenezcan. En la Figura 47 se muestra el comportamiento (en coordenadas polares) que sigue una partícula que constantemente fue afectada por la presencia de una observación. En la mayor cantidad de casos (más allá de la 15^o iteración) es posible notar que el módulo es creciente, implicando que la partícula se está alejando de su posición de origen, y con una orientación más o menos fija. Esta tendencia hace factible la estimación de su próxima posición mediante RLS.

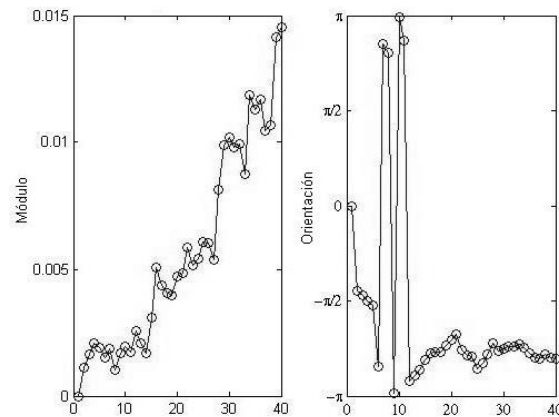


Figura 47: Ejemplo de trayectoria en coordenadas polares de una partícula que es afectada por la observación mayoritariamente. Se aprecia un claro aumento en la distancia entre la posición inicial y final de la partícula en cada iteración, indicando que se está alejando de su lugar de inicio para “explorar” otra zona del mapa en la que ocurren las observaciones recientes, mostrándose además una clara convergencia de la orientación del movimiento.

Revisando el caso contrario, es decir, aquellas partículas que mayoritariamente solo fueron afectadas por el ruido de proceso, se pueden apreciar comportamientos como el presentado en la Figura 48. En ella se aprecia que la distancia a la posición original crece y decrece a medida que se avanza en el tiempo, y a su vez la orientación del movimiento posee grandes variaciones, incluso con cambios excesivamente bruscos (iteración 10). Si bien con un filtro RLS es posible realizar una estimación de las posiciones futuras de esta partícula, significa una sobrecarga del procesamiento que no entrega mayor información, por lo que se omite con el fin de optimizar recursos y tiempo (recordando por supuesto que esta es una implementación en línea y el tiempo de procesamiento es una variable muy importante a considerar).

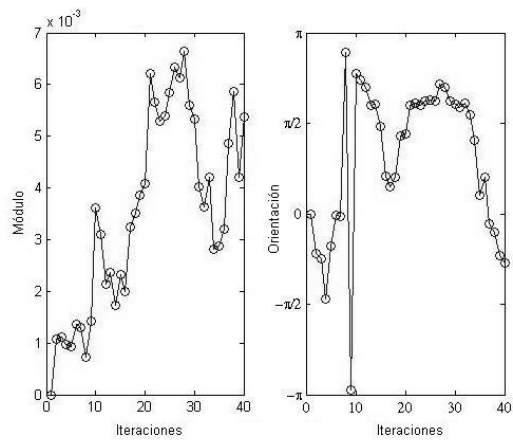


Figura 48: Ejemplo de trayectoria en coordenadas polares de una partícula que es afectada por el ruido de proceso mayoritariamente. La distancia entre la posición inicial y final en cada iteración se mantiene relativamente constante (notar orden de magnitud), creciendo y decreciendo con cada nueva observación. Esto indica que se mantiene cerca de su posición inicial, sin apreciarse tampoco tendencias en la orientación del movimiento.

Siguiendo la metodología propuesta, para cada partícula se obtendrá una predicción de su posición, y puede entonces aproximarse nuevamente una distribución normal bivariada en ese punto. En la Figura 49 se presenta un ejemplo del movimiento predicho por el filtro RLS para una partícula cualquiera: los movimientos anteriores de la partícula son utilizados para predecir una posición futura de esta, representado por la flecha negra en el acercamiento del plano. Es necesario mencionar que la última observación disponible no se utiliza en esta etapa, pero se ha colocado para entender que el movimiento de la partícula sigue siendo afectado por su presencia debido al uso del filtro.

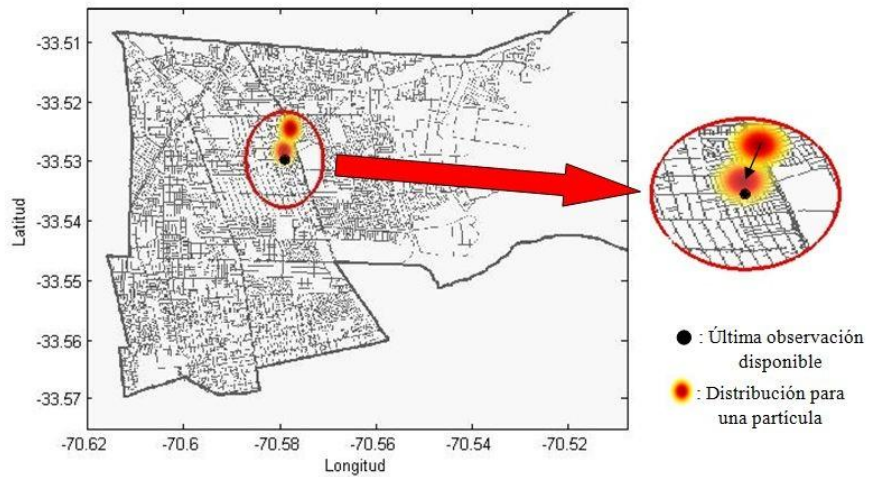


Figura 49: Ejemplo de la predicción a un paso mediante RLS para una partícula cualquiera. Considerando los movimientos previos (no mostrados en la imagen) y la última observación disponible se obtiene una predicción de la posición futura.

Al formar un nuevo modelo por GMM centrada en la predicción de la posición de cada partícula, se obtiene una predicción de la distribución espacial del riesgo, considerando la atenuación del movimiento y el aumento en la incertidumbre.

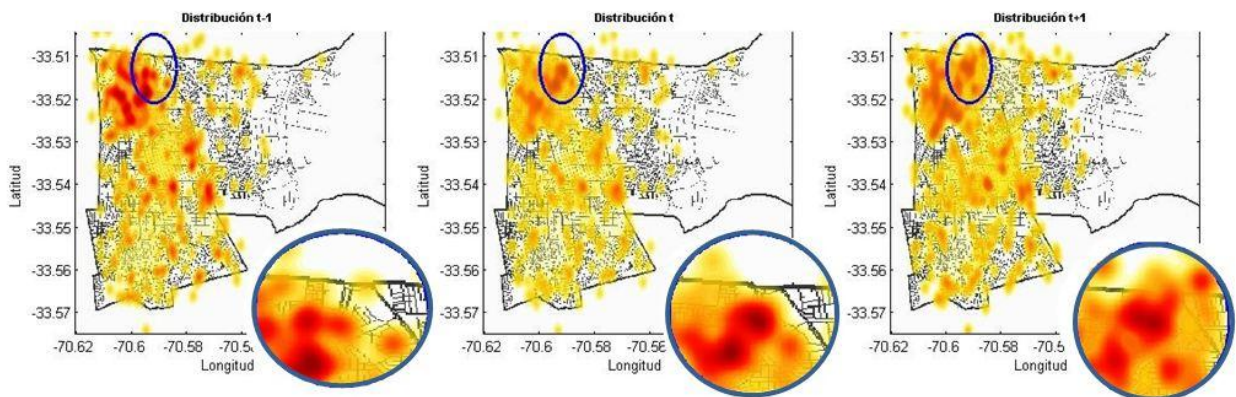


Figura 50: Ejemplo de la evolución temporal del modelo y predicción a un paso. Es posible apreciar en el acercamiento de cada imagen el movimiento de las partículas ubicadas dentro del umbral de acción de la última observación adquirida, apreciando una acumulación de estas y con ello un aumento del riesgo en esa zona. No son notorios los efectos de la atenuación del movimiento ni del aumento de la incertidumbre.

En la Figura 50 se presenta un ejemplo de la evolución temporal del modelo generado gracias a la última observación disponible y la predicción a un paso obtenida. En cada una de las distribuciones presentadas se entrega además un acercamiento a la zona donde efectivamente se

lleva a cabo el filtro RLS, pudiendo apreciarse que las partículas tienden a acercarse a dicha zona. Es ligeramente apreciable también el aumento en la incertidumbre, pero queda más claro en la Figura 51.

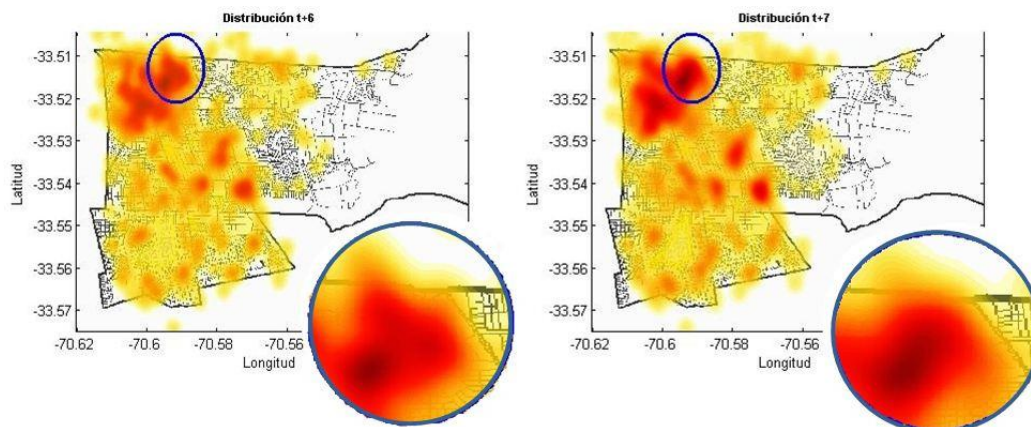


Figura 51: Ejemplo de la predicción a 6 y 7 pasos. Aquí ya es posible apreciar los efectos de la atenuación del movimiento y aumento de la incertidumbre, que se traducen en cambios poco notorios en la distribución del riesgo en la zona de interés.

En la Figura 51 se presentan las predicciones a 6 y 7 pasos, con el respectivo acercamiento a la zona de mayor relevancia. En ella es posible apreciar que la distribución ya no cambia en gran medida debido a la atenuación del movimiento de las partículas, y a la vez el aumento en la incertidumbre hace que el modelo se mantenga relativamente estable. Por esta razón se considera entonces el uso de un filtro RLS de orden 7, para incorporar la mayor cantidad de información al modelo pero sin predecir cuando ya la incertidumbre es demasiado alta. Esto previene el malgaste de recursos computacionales, que como se dijo anteriormente, dado que se trata de un proceso en línea es vital la rapidez de procesamiento.

Utilizando el mismo criterio de evaluación descrito en la Sección 2.10, es posible evaluar los modelos generados para las predicciones, considerando esta vez el conjunto de validación para evitar cualquier sesgo que puede incluirse involuntariamente con las observaciones utilizadas para definir el criterio de evaluación. En la Figura 52 se presentan los indicadores AUC para los modelos que se generan en cada predicción, considerando solo hasta la predicción a 7 pasos dado que más allá de este valor solo se aumenta la incertidumbre (ver Figura 51).

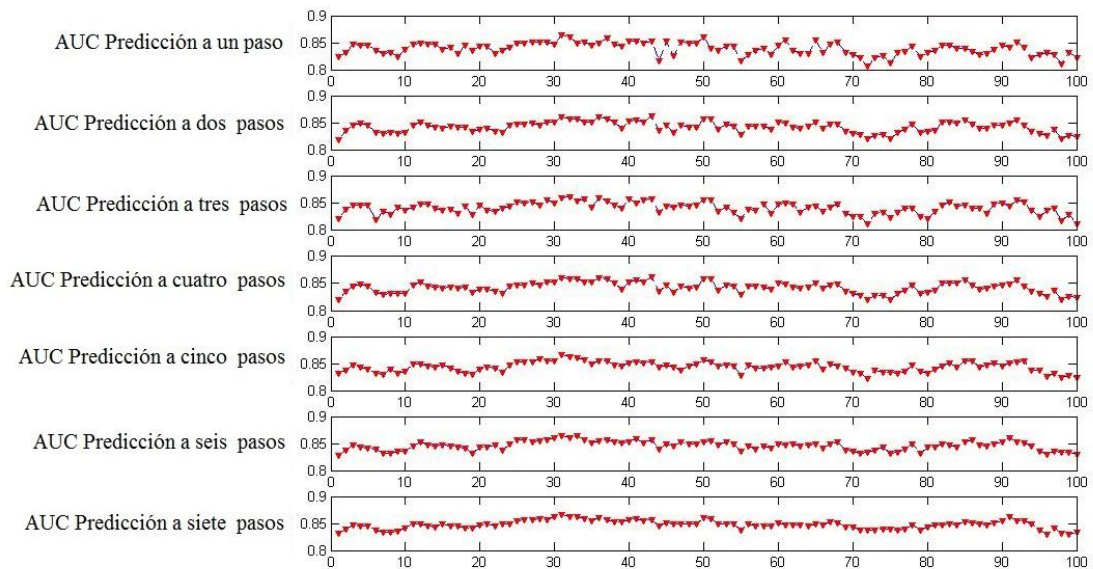


Figura 52: Cambios secuenciales del indicador AUC para las predicciones generadas mediante la metodología propuesta. A medida que se realiza predicción a más pasos se aprecia cada vez cambios más sutiles en el indicador entre un tiempo y otro, debido a los efectos de la atenuación y aumento de la incertidumbre. En ningún caso se obtienen valores inferiores a 0,8 indicado que en cada serie los modelos generados representan fielmente los incidentes futuros y por ende el riesgo en el área de interés.

Es posible apreciar que para todos los modelos generados mediante el algoritmo de evolución temporal y las predicciones correspondientes caracterizan de buena manera la probabilidad de ocurrencia de incidentes futuros. En la Tabla 3 se presentan los valores promedio y varianzas para cada predicción, corroborando los resultados expuestos al obtenerse como valor mínimo promedio un AUC de 0,8389.

Tabla 3: AUC para las predicciones del modelo probabilístico.

	Promedio [AUC]	Varianza [AUC ²]
Predicción a un paso	0,8389	$1,33 \cdot 10^{-4}$
Predicción a dos pasos	0,8397	$1,19 \cdot 10^{-4}$
Predicción a tres pasos	0,8398	$1,18 \cdot 10^{-4}$
Predicción a cuatro pasos	0,8418	$9,91 \cdot 10^{-5}$
Predicción a cinco pasos	0,8437	$8,48 \cdot 10^{-5}$
Predicción a seis pasos	0,8457	$7,23 \cdot 10^{-5}$
Predicción a siete pasos	0,8476	$6,3 \cdot 10^{-5}$

Se presentan en la Figura 53 la comparación entre las predicciones obtenidas para la distribución espacial de riesgo y los incidentes futuros de un conjunto de prueba. Al igual que antes, los nuevos eventos se aprecian cercanos o dentro de las zonas de alto riesgo.

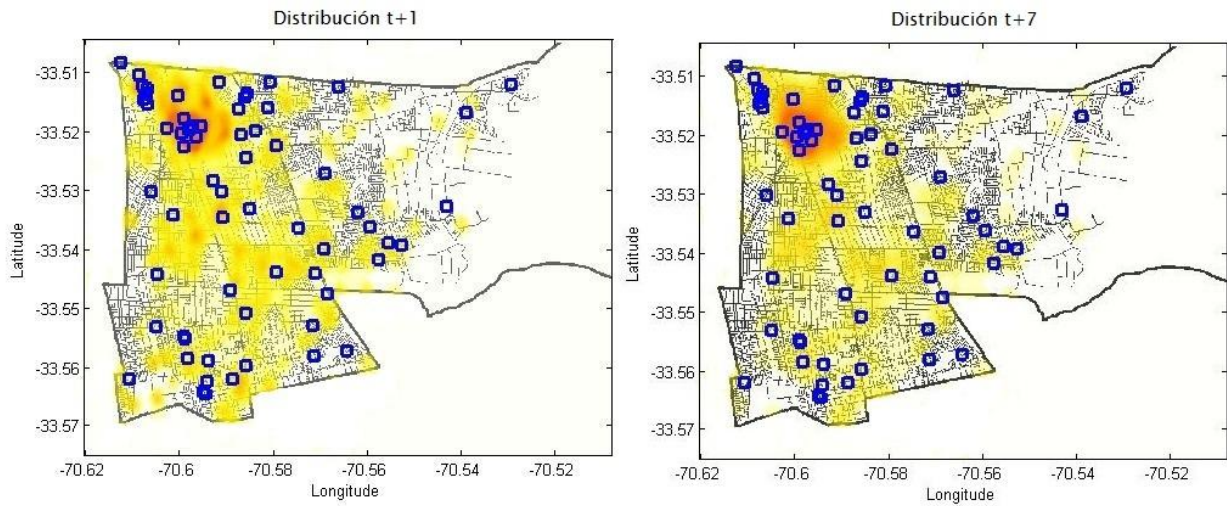


Figura 53: Esquema comparativo entre las predicciones de la distribución de riesgo y eventos futuros del conjunto de prueba. Nuevamente es posible apreciar la relación existente entre las zonas caracterizadas como “de alto riesgo” y la ocurrencia de eventos delictuales en ellas.

CAPÍTULO 5. CONCLUSIONES Y TRABAJO FUTURO

La policía tiene la necesidad de modelar patrones de riesgo asociados a la actividad delictual, con el objetivo de estudiar relaciones de causalidad y estimar dónde y cuándo un nuevo evento delictual puede ocurrir, siendo esta información de vital importancia a la hora de asignar recursos y mejorar la seguridad. En base a esto, se propuso como objetivo general de este trabajo el diseño de una metodología capaz de generar modelos espaciales de riesgo y caracterizar su evolución temporal con el fin de predecir dónde encontrar los focos delictivos en el futuro. Para lograr este propósito, se presentaron una serie de objetivos específicos que permitieron cumplir el objetivo descrito, pudiendo entonces mencionarse lo siguiente:

Teniendo presente que la ocurrencia de un delito se encuentra en la mayoría de las ocasiones ligado a la presencia de un foco delictivo, se diseña una metodología capaz de generar automáticamente distribuciones espaciales de riesgo, tomando como únicas entradas la ubicación de eventos delictuales y servicios, definidos dentro de un área e intervalo de tiempo determinados. La idea principal radica entonces en la obtención de estas distribuciones de riesgo pudiéndose ubicar a simple vista aquellas zonas en las que la probabilidad de ocurrencia de un delito es mayor (los focos delictivos) asistiendo así a las policías en la asignación de recursos en las zonas de mayor conflicto.

Para la evaluación de los modelos probabilísticos generados (distribuciones de riesgo), se utiliza un indicador similar al área bajo la curva (AUC), que relaciona el porcentaje de las áreas de alto riesgo y el porcentaje de eventos futuros que ocurren en dichas áreas de alto riesgo. Los modelos que se generan y evalúan en el caso de estudio (tanto en la etapa de modelación como en la evolución temporal y predicciones) corresponde a la utilización de toda la información disponible, obteniéndose un AUC superior a 0,83. Esto indica que las metodologías propuestas resultan satisfactorias en su capacidad de modelamiento, evolución y predicción respectivamente, principalmente al lograr que las áreas de alto riesgo sean pocas en comparación a las áreas totales de la zona de interés, y logrando también que un alto porcentaje de eventos futuros ocurran en estas zonas. El método de evolución temporal propuesto permite modificar el modelo probabilístico secuencialmente, incorporando nueva información a medida que esta se encuentra

disponible, y de acuerdo al AUC obtenido para cada iteración se puede concluir que la propuesta resulta satisfactoria no solo en la evolución del modelo sino también en su eficacia, al igual que el esquema de predicción de los focos de criminalidad.

No obstante el valor obtenido, para una implementación futura se hace necesario eliminar del análisis aquellas zonas que están fuera del área de interés delimitando correctamente la frontera de la zona de interés. En el caso de estudio, por ejemplo, se busca un modelo de la comuna, pero dado que no se dispone de coordenadas exactas que delimiten su contorno, se modela en toda el área rectangular que la rodea, dejando muchas zonas de la periferia involucradas en el análisis. Si mediante un software apropiado se lograra identificar estas regiones y no considerarlas en la evaluación, los valores del AUC se verán alterados al aumentar el porcentaje de áreas de alto riesgo para un umbral determinado. También, si esta metodología se quisiera escalar a más comunas, se debe buscar una forma de complementar la información de aquellas que compartan la frontera (por ejemplo un foco en una zona limítrofe de una comuna puede afectar de gran manera a la otra), pero en el caso contrario se deben buscar otras formas de lidiar con este asunto. Se propone trabajar con Kernel de Epanechnikov para la aproximación de las distribuciones en lugar de utilizar Gaussianas, permitiendo así acotar cada distribución a la zona de interés y evitando utilizar todo el plano.

Si bien el modelo que se genera y evalúa en el caso de estudio presentado en este Trabajo de Título considera, como se dijo antes, toda la información disponible, se estudiaron casos de modelación en que se considera solo un pequeño grupo de servicios, con la finalidad de crear modelos de riesgo asociado solo a la presencia de este. Por ejemplo, la presencia de un bazar frente a una estación de servicio altamente victimizada quedará representada dentro de una zona de alto riesgo, pero este no está asociado a la presencia del bazar sino a la estación de servicio. Así, para efectos estratégicos puede resultar útil conocer qué tipo de servicio es aquel que es fuente del riesgo, pudiéndose obtener estos modelos con la metodología propuesta. Continuando con el ejemplo, dos estaciones de servicio con ubicaciones diametralmente opuestas no necesariamente generaran la misma distribución de riesgo, pudiendo utilizarse también esta información para asociarla a otros parámetros del entorno. Otra modificación importante que permite realizar la metodología propuesta, es la creación de modelos que asignen importancia

relativa a cada servicio, simplemente modificando el peso que se le concede en la generación del modelo por mezcla de gaussianas. Por ejemplo, el efecto de un robo a un pequeño almacén no es el mismo que se produce en el robo a una gran tienda comercial, siendo que el delito puede ser el mismo. El primero produce un mayor daño tanto a los dueños como al entorno, pudiéndose entonces asignar mayor importancia a este tipo de servicios para evitar mayores efectos negativos en la comunidad.

En cuanto a la caracterización de la evolución temporal, el modelo dinámico propuesto utiliza parámetros que deben ser definidos de mejor manera en caso de una implementación real. Se propone la búsqueda un algoritmo que permita relacionar otros parámetros de diseño, como por ejemplo el ruido de proceso, el área de interés, entre otros, para obtener los parámetros del modelo dinámico matemáticamente. Durante el desarrollo de este trabajo se pudo apreciar que existe una relación entre el área de interés, la matriz de covarianza utilizada para definir el modelo dinámico y la covarianza del ruido de proceso, mas no se profundizó en este tema dado que un análisis más profundo solo cobra relevancia en caso de querer escalar el proyecto. El umbral de acción de la función de ajuste del modelo dinámico se mantiene fijo durante todas las iteraciones, y existen casos en los que esto no resulta favorable. Cuando una gran cantidad de eventos ocurren en una zona de alta densidad de partículas, muchas de estas se ven afectadas consecutivamente, lo que se traduce en una representación exagerada del riesgo y una convergencia exacerbada de muestras a dicha zona. Se considera posible reducir este efecto mediante la modificación de un cierto porcentaje de partículas por cada iteración (por ejemplo, el 10% de las partículas más cercanas a la nueva observación) y complementarlo con la definición del umbral que se utiliza en la propuesta inicial. Así, solo en el caso descrito se modificarían menos partículas, permitiendo que estas cumplan de mejor manera su labor exploratoria del mapa.

Finalmente, la cantidad de datos disponible solo hace posible hacer una estimación en línea que funciona a nivel táctico, es decir, en el corto plazo. Extendiendo la metodología propuesta a la modificación secuencia considerando un intervalo semanal o mensual, es posible obtener predicciones en estos mismos rangos de tiempo, permitiendo utilizar esta herramienta a niveles estratégicos, es decir, en el largo plazo.

BIBLIOGRAFÍA

- [1] Centro de Estudios Públicos. (2013, Octubre) www.cepchile.cl. [Online].
http://www.cepchile.cl/dms/archivo_5388_3457/encuestaCEP_sep-oct2013.pdf
- [2] Fundación Paz Ciudadana. (2013) www.pazciudadana.cl. [Online].
<http://www.pazciudadana.cl/publicacion/anuario-de-estadisticas-criminales-2012/>
- [3] Diputado Felipe Harboe B., "SEGURIDAD PÚBLICA - ¿dónde estamos y hacia dónde vamos?," 2012.
- [4] Subsecretaría de Prevención del Delito. (2001-2014)
<http://www.seguridadpublica.gov.cl/index.html>. [Online].
http://www.seguridadpublica.gov.cl/tasa_de_denuncias_y_detenciones.html
- [5] Instituto Libertad y Desarrollo. Delincuencia: La Percepción de la Gente. [Online].
http://www.institutolibertadydesarrollo.cl/lyd/controls/neochannels/neo_ch3801/deploy/encuesta%20delincuencia.pdf
- [6] CEAMOS. [Online]. <http://www.ceamos.cl/ceamos/>
- [7] Rachel Boba, *Crime Analysis and Crime Mapping*.: Sage Publications, Inc., 2005, ch. 2.
- [8] Rachel Boba, *Crime Analysis and Crime Mapping*.: Sage Publications, Inc., 2005, ch. 3.
- [9] John Eck, Spencer Chainey, James Cameron, Michael Leitner, and Ronald Wilson, "Mapping Crime: Understanding Hot Spots," U.S. Department of Justice, 2005.
- [10] Walter L. Perry, Brian McInnis, Carter C. Price, Susan C. Smith, and John S. Hollywood,.: RAND Corporation, 2013, ch. 1.
- [11] H. Liu and D. Brown, "Criminal incident prediction using a point-pattern-based density model," in *International journal of forecasting*, 2003, pp. 603-622.
- [12] Y. Xue and D. Brown, "Spatial analysis with preference specification of latent decision makers for criminal event prediction," in *Decision support systems*, 2006, pp. 560-573.
- [13] M.A Smith and D.E Brown, "Application of Discrete Choice Analysis to Attack Point Patterns," in *Information Systems and e-Business Management*, 2007, pp. 255-274.
- [14] D. Brown, J. Dalton, and H. Hoyle, "Spatial Forecast Methods for Terrorist Events in Urban Environments," in *Intelligence and security informatics: Second Symposium on*

Intelligence and Security Informatics, Tucson, AZ, 2004.

- [15] A. Rodrigues, P. Diggle, and R. Assuncao, "Semiparametric approach to point source modelling in epidemiology and criminology," in *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 2010, pp. 533-542.
- [16] C. Ivaha, H. Al-Madfai, G. Higgs, and A. Ware, "The Dynamic Spatial Disaggregation Approach: A Spatio-Temporal Modelling of Crime," in *Proceedings of the World Congress on Engineering, Volume 2, Citeseer 2007*.
- [17] Donald Brown and Xiaofeng Wang, "The spatio-temporal modeling for criminal incidents," in *Predictive Technology Laboratory*, University of Virginia, 2012.
- [18] Jean-Michel Marin, Kerrie Mengersen, and Christian Robert, *Bayesian Modelling and Inference on Mixtures of Distributions*.
- [19] D.A. Reynolds and R.C. Rose, "Robust Text-Independent Speaker Identification using Gaussian Mixture Speaker Models," in *IEEE Transactions on Acoustics, Speech, and Signal Processing* 3(1), 1995, pp. 72-83.
- [20] R.E. Kalman, "A new approach to linear filtering and prediction problems," in *Journal of Basic Engineering* 82(1), 1960, pp. 35-45.
- [21] B.D. Anderson and J.B. Moore, *Optimal Filtering*.: Dover Publications, 1979.
- [22] D. Simon, *Optimal State Estimation. Kalman, H-Infinity, and Nonlinear Approaches*, Wiley-Interscience, Ed., 2006.
- [23] B. Oksendal, *Stochastic Differential Equations: An Introduction with Applications*.: Springer, 2007.
- [24] F. Tobar, "Inferencia de la volatilidad de retornos financieros usando filtro de particulas," Universidad de Chile, Santiago, Chile, M.S. Thesis 2010.
- [25] P. Del Moral, *Peynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications, Probability and its Applications*. New York: Springer, 2004.
- [26] Sanjeev Arulampalam, Simon Maskell, Neil Gordon, and Tim Clapp, "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking," in *IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 50, NO. 2*, 2002, pp. 174-188.
- [27] Monson H. Hayes,.: Wiley, 1996, ch. 9.4: Recursive Least Squares.

- [28] G.D. González R., Identificación de Sistemas: Estimación Recursiva, 2013, Apuntes del curso EI7016: Optimización para el control de sistemas, FCFM, Universidad de Chile.
- [29] X. Wang and D. Brown, "The spatio-temporal generalized additive model for criminal incidents," in *Proceedings of the IEEE International Conference on Intelligence and Security Informatics*, Beijing, China, 2011.
- [30] T. Fawcett, "An introduction to ROC analysis," *Pattern recognition letters*, pp. 861-874, 2006.
- [31] Biblioteca del Congreso Nacional de Chile. (2013) [Online].
http://reportescomunales.bcn.cl/2013/index.php/La_Florida
- [32] ArcGIS. [Online]. <https://www.arcgis.com/features/>
- [33] OpenStreetMap. (2013) [Online].
<http://www.openstreetmap.org/relation/224590#map=15/-33.5151/-70.6019>
- [34] Real Academia Española. www.rae.es. [Online]. <http://lema.rae.es/drae/?val=cuadra>