Facultad de Medicina
Departamento de Medicina Preventiva y Salud Pública

Modelos anisótropos para el estudio de la mortalidad por cáncer en el entorno de focos contaminantes

Tesis Doctoral

Enrique Vidal Ocabo

Madrid 2011

Facultad de Medicina

Departamento de Medicina Preventiva y Salud Pública

Modelos anisótropos para el estudio de la mortalidad por cáncer en el entorno de focos contaminantes

Tesis Doctoral

Enrique Vidal Ocabo

Madrid 2011

Dirigido por: Gonzalo López-Abente y Roberto Pastor-Barriuso

D. Gonzalo López-Abente Ortega y D. Roberto Pastor Barriuso, Investigadores Titulares del Centro Nacional de Epidemiología del Instituto de Salud Carlos III

INFORMAN:

Que D. Enrique Vidal Ocabo ha realizado bajo su dirección el trabajo "Modelos anisótropos para el estudio de la mortalidad por cáncer en el entorno de focos contaminantes". Es un trabajo original, rigurosamente realizado, y es apto para ser defendido públicamente con el fin de obtener el grado de doctor.

Y para que así conste se firma el presente informe en Madrid a 9 de Mayo de 2011

Fdo: Gonzalo López-Abente Ortega Fdo.: Roberto Pastor Barriuso

Índice general

Índice de figuras

Índice de tablas

CAPÍTULO 1

Introducción

Las enfermedades no se distribuyen de manera homogénea en el espacio ni en el tiempo y, por supuesto, varían en función de las características de cada individuo. La epidemiología trata de estudiar esta distribución y explicar parte de su variabilidad, permitiendo entender los mecanismos etiológicos y tratando de mejorar el estado de salud general de la población. La rama de esta disciplina que se encarga de la variabilidad relacionada con la ubicación se conoce como epidemiología espacial y su cometido es investigar cómo la situación geográfica de las poblaciones condiciona su estado de salud. Los factores ambientales, como la contaminación, son algunos de los posibles mecanismos que influyen en esta relación.

Esta tesis trata de ampliar la metodología estadística espacial existente para el estudio epidemiológico de poblaciones residentes en entornos con focos contaminantes.

1.1. Epidemiología del cáncer

El cáncer es una enfermedad multifactorial de la que aún se desconoce parte de su etiología. De manera general, ésta puede separarse en dos componentes:

La primera corresponde con las características intrínsecas del individuo. Edad y sexo son los factores que más influyen en la aparición y desarrollo de la enfermedad. También los antecedentes genéticos están relacionados.

La otra componente es externa y engloba distintas categorías. Por una parte están las características socio-económicas y de estilos de vida, como la dieta, el consumo de alcohol, el sedentarismo o el tabaquismo. Existen también factores externos microbiológicos responsables de parte de los casos; Los virus de la hepatitis y del papiloma humano y las bacterias *Helicobacter pylori* son ejemplos de ellos. Por último hay que destacar la exposición ambiental a condiciones físicas (radiaciones ionizantes) y sustancias químicas que están implicadas en los distintos estadios del cáncer (Adami et al. 2008). Esta categoría incluye la exposición ocupacional.

Existe una gran diversidad de localizaciones y morfologías tumorales. Las más importantes están indicadas en la Tabla 1.1. La etiología varía mucho dependiendo de la localización y es, en general, resultado de la interacción de varios de los factores señalados. Por eso, el estudio del cáncer debe de abordarse teniendo en cuenta estas diferencias.

Respecto a la carga que produce en la población, en la Tabla 1.1 se comparan los datos de Europa y España para el año 2008 (Ferlay et al. 2010).

Las dos primeras columnas indican el tipo de cáncer, mediante el clasificación internacional de enfermedades (CIE 10) y su nombre. A continuación se presenta el número total de casos (incidentes o defunciones) y la tasa estandarizada por edad por cada 100.000 personas. Este ajuste se realiza tomando la población mundial como estándar.

		Incidencia en Europa		Mortalidad en Europa		Mortalidad en España	
CIE	**Tumor**	**Número**	**Tasa**	**Número**	**Tasa**	**Número**	**Tasa**
C00-C14	*Labio, C. Bucal y Faringe*	67764	8	26267	2.9	2113	5.53
C11	*Nasofaringe*	2980	0.4	1484	0.2	185	0.5
C15	*Esófago*	33013	3.5	28758	2.9	1781	4.57
C16	*Estómago*	83120	7.9	62128	5.6	5614	11.31
C18-C21	*Colorrectal*	334092	31.7	149159	12.6	13793	26.12
C22	*Hígado*	48219	4.7	45919	4.2	4519	9.4
C23-C24	*Vesícula Biliar*	23091	2	17318	1.4	1227	2.09
C25	*Páncreas*	69661	6.6	71116	6.5	5216	10.81
C32	*Laringe*	28344	3.4	12710	1.4	1546	3.94
C33-C34	*Pulmón*	289406	30.2	254031	25.2	20170	48.89
C43	*Melanoma*	69387	9.1	14243	1.5	875	2.01
C50	*Mama*	332670	77.1	89801	16.6	6105	12.79
C53	*Cuello del Útero*	31038	9	13430	3	612	1.51
C54	*Cuerpo del Útero*	56979	11.7	13155	2	913	1.58
C56,C57	*Ovario*	44728	9.7	28924	5.2	1975	4.09
C61	*Próstata*	323790	69.5	71027	12.1	5458	9.42
C62	*Testículo*	15255	6	931	0.3	49	0.16
C64-C66,C68	*Riñón*	73171	8	31322	2.9	2129	4.47
C67	*Vejiga*	107419	10	38641	3	4701	8.74
C70,C71,C72	*Encéfalo y SNC*	41059	5.6	31484	3.8	2599	6.79
C73	*Tiroides*	33599	5	3581	0.3	305	0.6
C81	*Linfoma de Hodgkin*	11777	2.1	2631	0.3	240	0.61
C82-C85,C96	*Linfoma no Hodgkin*	74162	8.3	31371	2.8	2504	5.03
C90	*Mieloma*	31792	3	20802	1.7	1630	2.9
C91-C95	*Leucemias*	59375	7.1	40551	3.8	3032	6.42
C00-C97	*Tumores Malignos*	2444597	264.3	1234303	114.7	100439	212.51

Tabla 1.1: Número de casos y tasas ajustadas (pobración mundial) por cada 100.000 personas

En resumen: A lo largo del año 2008 se registraron 100.349 muertes por cáncer en España (Área de Epidemiología Ambiental y cáncer Centro Nacional de Epidemiología ISCIII 2008). En términos relativos, el cáncer supuso el 20 y 32 % sobre el total de defunciones en mujeres y hombres respectivamente. En Europa, 3.2 millones de personas enfermaron y más de 1.7 millones murieron de cáncer ese mismo año (Ferlay et al. 2010).

Hay varias consideraciones que hacen que la mortalidad sea uno de los indicadores más frecuentes a la hora de estudiar el cáncer (Adami et al. 2008). Su reducción es un objetivo usual en los programas de control. La defunción es un evento inequívoco, su registro sistemático está extendido en la mayoría de los países y es accesible, a nivel poblacional, en la mayoría de los casos. También presenta algunas limitaciones, ya que está condicionada no sólo por la incidencia de la enfermedad sino también por su supervivencia, así como por la calidad de los certificados de defunción.

Por todo lo expuesto, la justificación del estudio del cáncer es doble: Todavía quedan aspectos de su etiología por descubrir y afecta a un gran número de personas.

1.2. Exposición ambiental a focos contaminantes

Una de las posibles causas ambientales del cáncer es el conjunto de sustancias tóxicas emitidas de forma constante al ambiente por algunas insta-

laciones industriales. Algunas de las sustancias han sido identificadas por la IARC (*International Agency for Research on Cancer*) con distintos grados de potencial carcinogénico.

A nivel internacional se han descrito asociaciones entre la enfermedad y residir en las proximidades de industrias que emiten este tipo de sustancias. Gottlieb et al. (1982) describieron una asociación de la incidencia de cáncer de pulmón con la proximidad de industria metalúrgica, complejos industriales y otras fuentes de emisión. Parodi et al. (2005) y Edwards et al. (2006) señalaron su relación con la cercanía a coquerías e industrias pesadas respectivamente. Existen evidencias de que la proximidad a áreas industriales aumenta la frecuencia de leucemias y linfomas (Benedetti et al. 2001; Linos et al. 1991; Johnson et al. 2003; Sans et al. 1995). También hay estudios que no han encontrado asociación con la proximidad de industrias contaminantes (Michelozzi et al. 1998; Pekkanen et al. 1995; Elliott et al. 1992).

En España el número de publicaciones de este tipo es menor, debido en parte a la falta o dificultad de acceso a la información, aunque sí existen algunos trabajos concretos que estudiaron distintos tipos de industrias: Incineradoras (Gonzalez et al. 2000), industria nuclear (López-Abente et al. 2001; Silva-Mato et al. 2003), plantas electroquímicas (Ozalla et al. 2002; Sunyer et al. 2002; Grimalt et al. 1994) e instalaciones que usaban asbesto (Magnani et al. 2000).

Por otra parte, el patrón de mortalidad municipal mostrado por algunos tumores en el Atlas municipal de mortalidad por cáncer en España (López-

Abente et al. 2007) podría indicar la existencia de causas ambientales. En concreto, su coincidencia con áreas muy industrializadas sugiere que exposiciones derivadas de la actividad industrial podrían estar condicionando la salud de su entorno.

Tanto las autoridades como los medios de comunicación y el público general sienten interés en conocer si la presencia de focos potencialmente contaminantes influye en la salud de las poblaciones que residen en su entorno. La escasez y elevado coste de mediciones individuales de niveles de contaminación así como el gran número y variabilidad de escenarios a considerar dificultan un abordaje pormenorizado y preciso. Por lo tanto son necesarias herramientas que permitan realizar estudios exploratorios sencillos y rápidos; que hagan uso de fuentes de datos disponibles y se puedan llevar a cabo de manera sistemática sin grandes requerimientos de recursos personales, materiales ni temporales.

1.3. Estudios ecológicos de datos agregados

La epidemiología espacial estudia la relación de la variabilidad de las enfermedades con la ubicación geográfica (Elliott et al. 2001). Por una parte trata de describir la situación existente, mediante la cartografía de enfermedades y la detección de "clusters" (Lawson et al. 1999, 2003; Waller and Gotway 2004; Bivand et al. 2008). Por otra, estudia su relación con potenciales factores de riesgo como indicadores sociales, demográficos, sanitarios y económicos o posibles focos contaminantes (Elliott and Wartenberg 2004).

El concepto sobre el que se construye la metodología epidemiológica es el de población humana (Rothman et al. 2008; Szklo and Nieto 2007). La base del estudio es la agrupación de personas, atendiendo a los criterios que se consideren relevantes, y su posterior clasificación. En el caso de la estadística espacial aplicada a estudios exploratorios, la agrupación se realiza teniendo en cuenta la localización geográfica, comparando las características de los diferentes conjuntos de personas incluidos en el estudio.

Una manera eficaz de realizar este tipo de análisis exploratorios es un estudio ecológico (Waller and Gotway 2004). En él se agrega la información individual en niveles jerárquicos superiores que definen las áreas de estudio (por ejemplo secciones censales, municipios, provincias). Este tipo de indicadores agregados suele ser de dominio público o de fácil acceso. No obstante la agregación de casos puede introducir sesgos. Uno de los más importantes es la llamada "falacia ecológica" descrita por Greenland and Morgenstern (1989). Se debe a que los resultados obtenidos a un nivel de agregación dado (por ejemplo, municipal) no tienen por qué ser directamente trasladables a niveles inferiores (individuos).

Ante la dificultad de obtener medidas precisas de dosis efectivas individuales o promedio para las áreas de estudio es frecuente utilizar la localización geográfica como medida aproximada de la exposición. En los escenarios con focos contaminantes lo más común es usar la distancia al foco. En el capítulo 1.4 se presenta una revisión de los métodos más utilizados.

Para medir de manera cuantitativa el estado de enfermedad de una población es necesario recurrir a indicadores colectivos que recojan esta infor-

mación. Los más utilizados (aunque no lo únicos) son las tasas de incidencia y mortalidad. Se definen como el cociente entre el número de personas enfermas o fallecidas por una determinada causa en una población a lo largo de un periodo de tiempo y el total de personas-tiempo a riesgo de padecer dicha situación. Estas tasas suelen presentarse desagregadas por sexo. Además, a la hora de comparar las tasas entre dos o más poblaciones, hay que tener en cuenta que las diferencias entre las distribuciones de edad pueden condicionar el resultado. Por eso es necesario ajustar estos indicadores teniendo en cuenta la estructura etaria de la población, de manera que sean más comparables. La estandarización de tasas permite un abordaje de esta situación. Se trata de calcular un promedio ponderado de la tasas específicas por grupos de edad. En la práctica suele emplearse el método indirecto de estandarización, ya que es más estable que el método directo en poblaciones con un reducido número de eventos y no requiere conocer las tasas especificas por edad de cada población. En concreto, se utiliza como medida de asociación la razón de mortalidad estandarizada (RME), que corresponde al cociente entre el número observado de casos en la población a estudio y el número de casos que cabría esperar al aplicar las tasas especificas por edad de la población de referencia a la distribución por edad de la población a estudio.

Una vez se dispone de la información al nivel de agregación deseado de los indicadores de salud y de los posibles factores de riesgo se estudia su asociación. Esto se suele llevar a cabo mediante un análisis de regresión, circunstancia que da nombre a los estudios de regresión ecológica.

1.4. Antecedentes metodológicos

Las metodologías para abordar el estudio de los indicadores de salud en el entorno de focos contaminantes se pueden clasificar atendiendo a la modelización de la relación de dependencia entre ubicación espacial y riesgo (Elliott et al. 2001). Por una parte, están los test no paramétricos de asociación que no requieren asunciones con respecto a las características de esta relación. Por otra, los abordajes paramétricos que estudian la dependencia espacial del riesgo imponiendo una estructura específica, con distintos grados de flexibilidad. Dentro de este último conjunto se encuentran los modelos de regresión.

En este capítulo se expone una pequeña revisión de los métodos más utilizados, destacando sus propiedades, ventajas y limitaciones. Para ilustrar gráficamente alguno de ellos se va a utilizar como ejemplo una industria de fabricación de armas en Oviedo. La elección de este escenario responde a criterios puramente académicos, pues presenta una distribución espacial que facilita la comprensión de los modelos presentados. En la Figura 1.1 se muestra la disposición espacial de las áreas de estudio en el entorno de un foco emisor (*). El tamaño de los puntos es proporcional a la RME.

Antes de comenzar con la revisión de modelos basados en la localización geográfica, que son la mayoría de los utilizados en este campo, hay que reseñar que existen otras alternativas basadas en estimaciones del nivel de contaminantes en cada población. Un ejemplo de ellas son los modelos de dispersión atmosférica. Esta metodología en primer lugar estima los niveles

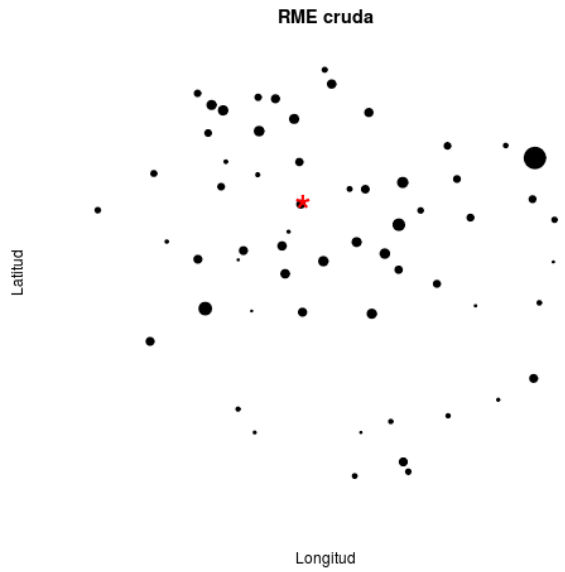


Figura 1.1: Distribución espacial de los municipios alrededor del foco (asterisco rojo). El tamaño de los puntos es proporcional a la RME

de inmisión de contaminantes en el entorno de los focos emisores a partir de información relativa a las características particulares de cada uno (tasa de emisión, altura de la chimenea, etc ...) y su entorno físico (orografía, meteorología, etc ...), para posteriormente utilizar esta información en un análisis de regresión. Se pueden encontrar aplicaciones de este tipo de aproximación en Viel et al. (2008) y Goria et al. (2009). La minuciosidad y especificidad de la información utilizada permiten mejores caracterizaciones de la exposición, pero restringen su factibilidad; no siempre se dispone de detalles a cerca del funcionamiento de los focos contaminantes ni de las todas las características físicas de cada entorno estudiado. Esto es una limitación importante a la hora de realizar un análisis exploratorio sistemático.

Modelos basados en la distancia al foco

Aunque existen abordajes en los que se tiene en cuenta la posición, la información más utilizada para estudiar los entornos de focos contaminantes es la distancia al emisor. Si se sospecha que el deterioro de la salud pueda estar relacionado con los contaminantes emitidos por el foco y si se puede asumir que la difusión de dichos contaminantes depende de la distancia, una asociación positiva entre la proximidad al foco y un empeoramiento de los indicadores de salud apunta al foco emisor como causante posible del deterioro. En la Figura 1.2 se ha representado la RME en función de la distancia al foco emisor para cada una de las áreas del ejemplo. También aparece la RME (con su IC 95 %) agregada por distancia.

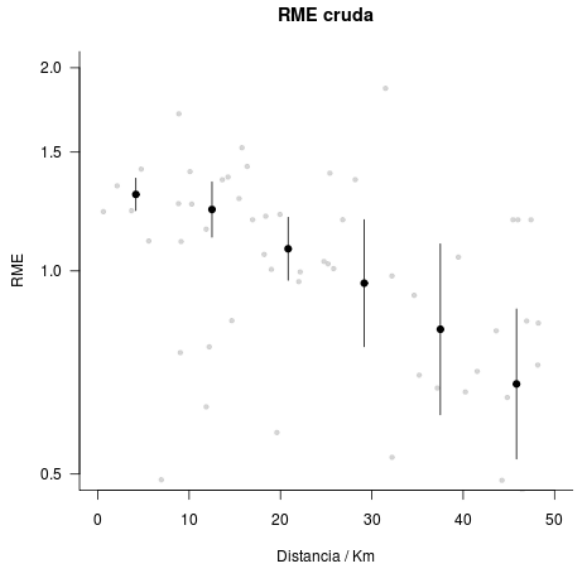


Figura 1.2: RME e IC95 % en función de la distancia al emisor en el entorno de un foco contaminante

Los modelos revisados en esta sección se pueden clasificar en dos grandes

grupos. Por una parte, los tests no paramétricos, que no asumen ninguna forma subyacente de la asociación del riesgo con la distancia. Por otra, los modelos de regresión.

Tests no paramétricos de asociación

La metodología no paramétrica permite contrastar la asociación entre la posición y el riesgo prescindiendo de asunciones concretas acerca de su forma.

Partiendo de métodos para la detección de clusters de enfermedades (Openshaw et al. 1988), Besag and Newell (1991) y Newell and Besag (1996) propusieron un test no paramétrico para comprobar la existencia del riesgo asociado a un foco. La hipótesis que se contrasta es que los casos se distribuyen aleatoriamente en la población. Para ello, se divide la zona de estudio en áreas disjuntas y ordenadas con respecto a su distancia al foco. El estadístico de contraste es el número mínimo de áreas que es necesario agrupar para acumular un numero determinado de casos. El nivel de significatividad se obtiene suponiendo que los casos siguen una distribución de Poisson con media igual al número esperado de casos en base a las tasas estandarizadas según una población común de referencia. El control de las posibles variables confusoras se incluye en el cálculo de los casos esperados.

Esta aproximación es fácil y rápida de implementar. Sin embargo, la elección arbitraria del número de casos a acumular condiciona los resultados (Waller and Lawson 1995), limitando su uso. Además, al definir y ordenar

las áreas con respecto a la distancia al foco se está descartando de manera implícita una posible relación del riesgo con la dirección.

Una alternativa ampliamente extendida es el test introducido por Stone (1988). En él también se subdivide la zona de estudio en áreas en las que se supone que el número de casos sigue la distribución de Poisson. A partir de ahí se trata de comprobar si las tasas de cada una de ellas se distribuyen de manera no creciente con respecto a algún indicador de exposición que permita ordenarlas (típicamente la distancia). Mediante simulaciones de Monte Carlo se consigue estimar la significatividad del contraste.

Este test puede extenderse adaptándose a diversas situaciones, como escenarios con más de un foco (Shaddick and Elliott 1996) o permitir ajustes por covariables (Morton-Jones et al. 1999). En este caso se hace uso de modelos de regresión, con lo que estos test son un híbrido entre las dos categorías presentadas. Tampoco aquí se asume ninguna relación entre riesgo y distancia a priori. De todas maneras, la partición de las áreas de estudio se realiza en base a la distancia o bien a información suplementaria (exposición). En el primer caso se vuelve a obviar el posible componente direccional del riesgo y el segundo requiere información extra que no siempre está disponible.

Si se interpretan las particiones de la zona de estudio como estratos de exposición es posible generar una serie de tests semi paramétricos a partir de la diferencia entre el número de casos observados y esperados que se resuelven mediante una distribución chi cuadrado (Tarone 1982; Breslow et al. 1983). Además, se puede incluir un término de ponderación por algún tipo de indicador de la exposición, como por ejemplo, la distancia.

Existen más métodos no paramétricos como la familia de "Linear risk scores" propuesta por Bithell (1995) así como adaptaciones de metodologías heredades de la detección de clusters de enfermedades (Cuzick and Edwards 1990).

Todas las alternativas propuestas hasta el momento permiten contrastar la existencia de una asociación entre la ubicación de un foco contaminante y la enfermedad en su entorno sin asunciones respecto a la forma de la relación. Sin embargo, tienen tres limitaciones importantes:

1. Las categorizaciones son arbitrarias
2. Tienen una capacidad limitada de ajuste (a través de los valores esperados)
3. Ninguna permite cuantificar las estimaciones del riesgo

Modelos de regresión

Por contraposición a los métodos no paramétricos, en esta sección se presentan modelos en los que se establece una estructura determinada para estudiar la relación de dependencia espacial del riesgo. Esto permite solventar las limitaciones fundamentales de los tests no paramétricos.

El grado de flexibilidad de las asunciones que estas estructuras requieren es variable. En general, las aproximaciones que imponen condiciones rígidas tienden a ser fáciles de implementar e interpretar, pero pueden alejarse mucho de la realidad que se quiere modelizar. Por el contrario, los abordajes más flexibles suelen implicar una mayor dificultad a la hora de ponerlos

en práctica y/o extraer conclusiones. Los distintos grados de flexibilidad permiten establecer la clasificación que se presenta a continuación.

Modelos categóricos La manera más directa de estudiar la asociación espacial de un foco contaminante con su entorno es comparar el riesgo de una zona expuesta con otra de referencia. A partir de una distancia especificada a priori se delimitan dos zonas: una cerca, que se supone expuesta, y otra lejos, supuesta de referencia (Breslow et al. 1993; Clayton and Hills 1993). La comparación de los indicadores entre las dos zonas produce una estimación del riesgo relativo en la zona expuesta con respecto a la de referencia.

La simplicidad de esta aproximación, en la que la información espacial se reduce a una variable de localización dicotómica, es la responsable de sus mayores ventajas así como de las limitaciones más importantes.

Por una parte, la inclusión de la variable creada en un modelo de regresión permite obtener una estimación del riesgo a partir de su coeficiente. Además, la interpretación es directa: se compara una zona expuesta con otra de referencia.

Por otro lado, se asume que cada una de las dos zonas presenta un riesgo homogéneo; no se contemplan variaciones del riesgo dentro de la zona expuesta (o de la de referencia). Así, por ejemplo, en el modelo de la Figura 1.3 se asume un riesgo constante tanto para los municipios situados a menos de 10 km del foco como para los situados a más de 10 km. En particular, el riesgo en los municipios próximos al foco es un 21 % (IC al 95 % 11-31 %) superior al de los municipios más distantes. También se introduce un

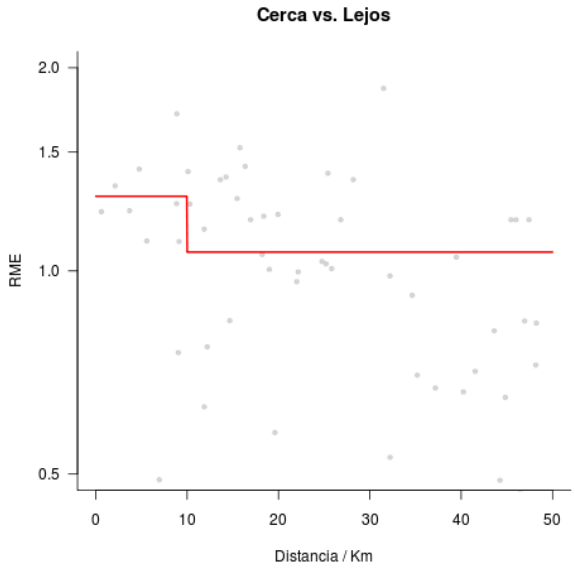


Figura 1.3: Comparación cerca vs. lejos. Los puntos se corresponden a la RME para cada municipio y la linea roja representa la curva de riesgo

elemento arbitrario (el alcance) que condiciona los resultados. Además, el modelo establece un cambio brusco en el riesgo (el "escalón" de la Figura 1.3) que suele tener escasa plausibilidad desde el punto de vista epidemiológico.

Una extensión de la comparación cerca vs. lejos que permite solventar la primera de las limitaciones comentada consiste en aumentar el número de zonas a tener en cuenta (Clayton and Hills 1993). Establecidas una serie de distancias ordenadas, se definen varios anillos concéntricos. Usualmente se elige el anillo exterior como zona de referencia y se procede de manera similar al análisis cerca vs. lejos, estimando el riesgo relativo de cada uno de los anillos con respecto al de referencia.

Como en el abordaje anterior, la estimación de los indicadores de riesgo se realiza mediante una regresión, en este caso incluyendo variables indica-

doras para cada uno de los anillos a excepción del de referencia. Las estimaciones (e IC al 95 %) de los riesgos relativos de las sucesivas categorías de la Figura 1.4 son 1.19 (1.22-1.36), 0.93 (0.84-1.02), 0.73 (0.61-0.87), 0.70 (0.56-0.89) y 0.53 (0.41-0.68) respectivamente.

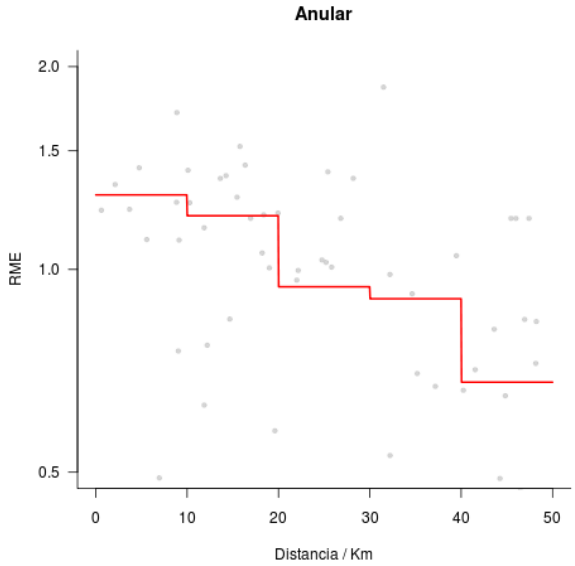


Figura 1.4: Comparación anular. Los puntos se corresponden a la RME para cada municipio y la linea roja representa la curva de riesgo

Se puede aprovechar la información contenida en la variable anular para investigar la existencia de una tendencia del riesgo con la distancia. Para ello, basta incluir en el modelo de regresión una variable ordinal continua con la distancia promedio de cada anillo, cuyo valor P asociado permite contrastar la existencia de una componente lineal en la tendencia del riesgo (López-Abente et al. (1999) sirve como ejemplo). Además, si el modelo anular categórico presenta evidencias que permitan justificar su uso frente al anular continuo, se puede suponer que existe una relación espacial más

compleja (no monótona).

Si bien esta aproximación permite modelizar la relación distancia-riesgo de manera más flexible que la anterior, evaluando no sólo asociación sino también tendencia, adolece aún más del mismo problema de arbitrariedad, ya que ahora son varias las distancias escogidas a priori para definir las categorías que resumen la información espacial. Además, si el número de categorías es excesivo (típicamente se seleccionan entre 3 y 5 categorías), habrá pocos municipios en cada categoría, lo que ocasionará una mayor imprecisión en las estimaciones que podría distorsionar la tendencia real subyacente. El problema de los cambios abruptos ("escalones") también continua presente.

Se puede mejorar la aproximación cerca vs. lejos incluyendo un parámetro que controle la distancia que delimita las zonas. El resultado es un modelo que estima no sólo el riesgo relativo de la zona más expuesta (y más cercana al foco) si no también su alcance (Figura 1.5). En el ejemplo se obtiene un aumento del 50 % asociado a los municipios que distan menos de 24 Km, que contrasta con el obtenido al establecer de manera arbitraria el alcance en 10 Km (21 % de aumento del riesgo).

Las ventajas y limitaciones de esa aproximación son las mismas que las de la comparación cerca vs. lejos. Sin embargo, se evita la arbitrariedad de la definición de las zonas. Por contra, se añade complejidad a la hora de estimar ya no sólo los valores de los parámetros, si no, sobre todo, su variabilidad.

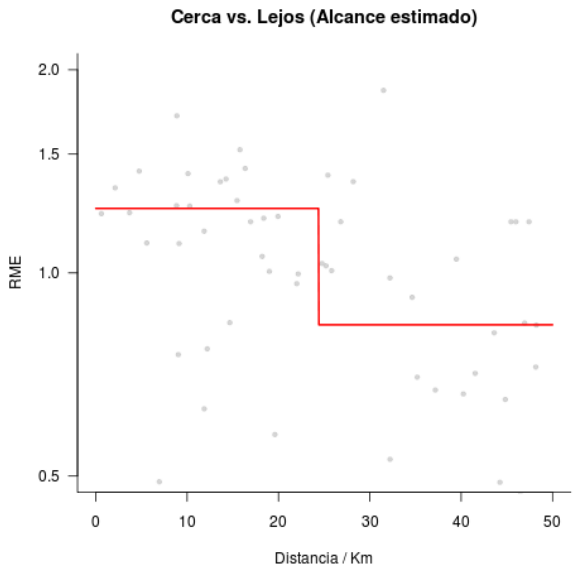


Figura 1.5: Modelo cerca vs. lejos con estimación del alcance. Los puntos se corresponden a la RME para cada municipio y la linea roja representa la curva de riesgo

Modelos lineales, logarítmicos y polinomiales Mediante una dependencia lineal o polinómica del riesgo con los factores espaciales se obtienen modelizaciones de implementación directa a partir de los procedimientos estándar de regresión linear generalizada (McCullagh and Nelder 1989). Básicamente se trata de incluir potencias de distintos ordenes de la distancia como términos en la regresión.

Dentro de este tipo de modelos, la manera más simple de abordar la modelización del riesgo espacial es un modelo lineal donde se introduce la variable distancia directamente en el modelo de regresión. El coeficiente obtenido relaciona cada incremento constante en la distancia con un mismo cambio porcentual del riesgo. Así, en el modelo lineal de la Figura 1.6 se estima que por cada incremento de 10 km en la distancia el riesgo disminuye

un 12 %, con un IC al 95 % entre 8 y 15 %.

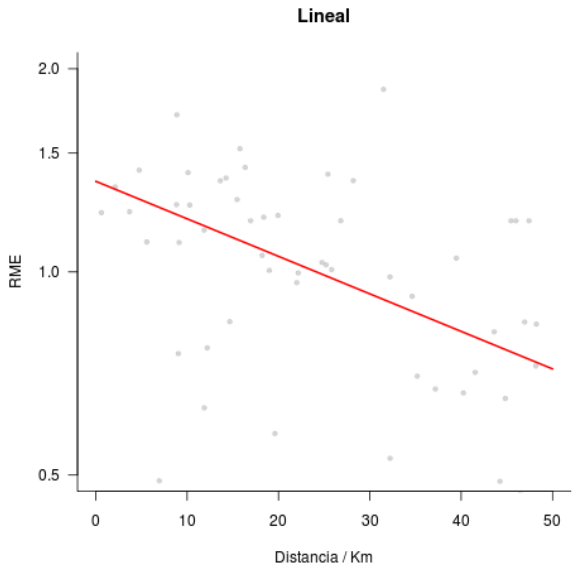


Figura 1.6: Modelo lineal. Los puntos se corresponden a la RME para cada municipio y la linea roja representa la curva de riesgo

La estimación y la interpretación del indicador en cuestión resultan directas e intuitivas. Este análisis no requiere ninguna categorización arbitraria a priori. Las unidades en las que se expresa la distancia condicionan el valor de la estimación puntual del incremento porcentual del riesgo, pero no el de su grado de significación estadística (valor P asociado).

Este modelo impone una proporcionalidad constante del riesgo con la distancia (relación lineal). Así, en la Figura 1.6, se asume un mismo cambio porcentual del riesgo entre 0 y 10 km de distancia al foco que entre 40 y 50 km, aun cuando en la práctica cabría esperar una atenuación del riesgo relativo conforme aumenta la distancia (Diggle and Elliott 1995).

También se pueden utilizar modelos logarítmicos, que presentan una

interpretación de los coeficientes muy sencilla: Cualquier incremento relativo de la distancia produce el mismo cambio porcentual en el riesgo. En la Figura 1.7 se observa la tendencia resultante de un modelo con el logaritmo de la distancia. El riesgo relativo se reduce un 15 % (IC 95 % 9-21) cada vez que se dobla la distancia al foco. Aunque el riesgo se atenúa con la distancia, éste se hace infinito en las proximidades del foco.

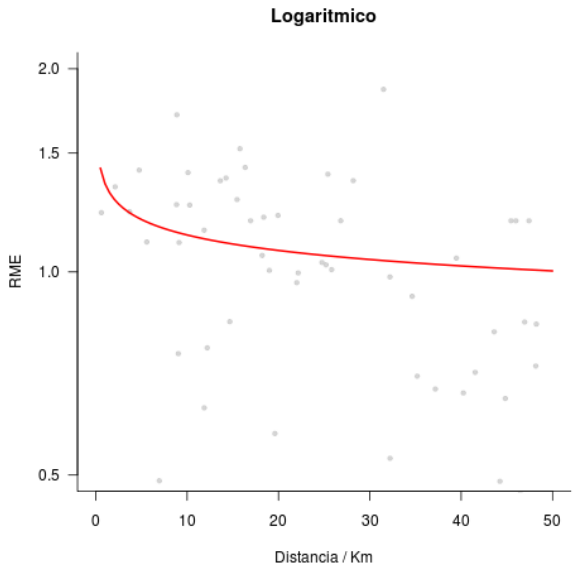


Figura 1.7: Modelo logarítmico. Los puntos se corresponden a la RME para cada municipio y la linea roja representa la curva de riesgo

En un modelo de regresión se pueden incluir combinaciones de distintas potencias enteras y fraccionarias de la distancia dando lugar a modelos polinomiales fraccionarios (Greenland 1995). La principal ventaja es su flexibilidad para amoldarse a cualquier relación subyacente: sin embargo, no pueden resumirse en una única medida de asociación. Se trata de modelos diseñados para su representación gráfica (en esto, muy similares a los no

paramétricos).

Si se introduce la inversa de la distancia en la regresión, se obtiene un caso particular de estos modelos (Figura 1.8). Se consigue solventar el problema a grandes distancias del foco: el riesgo decae asintóticamente. Sin embargo, el riesgo relativo en el foco se hace infinito, con lo que los resultados están muy influenciados por la proximidad del área más cercana al foco.

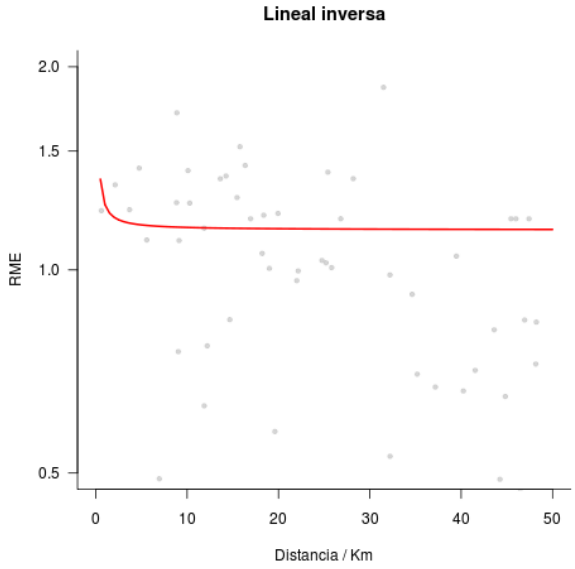


Figura 1.8: Modelo lineal inverso. Los puntos se corresponden a la RME para cada municipio y la linea roja representa la curva de riesgo

Un ejemplo de estos modelos se puede observar en la Figura 1.9, en donde se han introducido los términos de orden -1, -.5, .5, 1 y 2 de la variable distancia.

Así se obtiene una parametrización muy flexible, aunque existe cierto grado de arbitrariedad a la hora de escoger el orden del polinomio. También

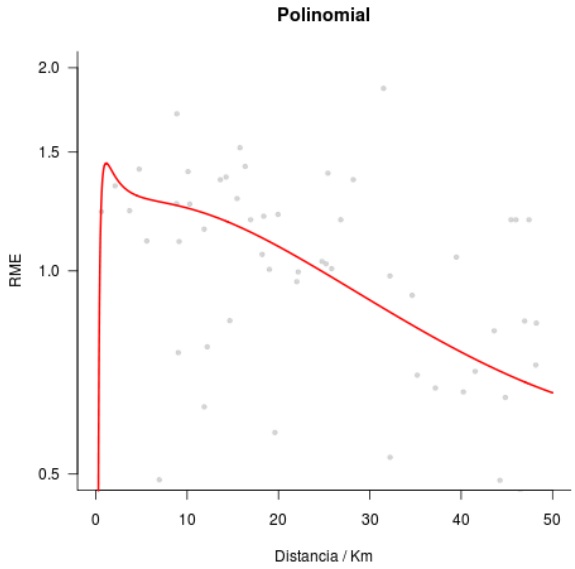


Figura 1.9: Modelo polinomial. Los puntos se corresponden a la RME para cada municipio y la linea roja representa la curva de riesgo

se puede dar la situación en la que la curva estimada presente formas complicadas, poco plausibles desde el punto de vista epidemiológico. Además, la inclusión de la distancia de manera recurrente en diversos órdenes polinomiales dificulta la interpretación de los coeficientes asociados. En particular, el modelo podría sufrir de problemas de multicolinealidad ya que las potencias de la distancia tienden a estar muy correlacionadas. No obstante, estos problemas pueden aliviarse notablemente centrando la variable distancia o utilizando términos polinomiales ortogonales.

Todos los modelos propuestos en este apartado pueden implementarse en un modelo de regresión que es lineal en los parámetros. Por ello, la estimación y posterior inferencia sobre ellos, que permitirá contrastar la existencia de una componente espacial del riesgo y cuantificarla, se realiza

mediante procedimientos estándar (McCullagh and Nelder 1989) y no reviste mayor complicación.

Modelos no lineales Las parametrizaciones hasta ahora expuestas permiten modelizar gran variedad de escenarios, pero se puede conseguir mayor flexibilidad mediante modelos no lineales. Se trata de incluir términos en la regresión que establecen relaciones no lineales con los parámetros. De esta manera se gana en plausibilidad epidemiológica, a costa de una mayor complejidad en la implementación, estimación e inferencia. En el trabajo de Diggle et al. (1997) se introducen una serie de dependencias funcionales del riesgo con la distancia que se exponen a continuación.

Se puede parametrizar un modelo no lineal con una disminución exponencial cuadrática (Figura 1.10). Con esto se consigue una distribución del riesgo finita en todo el rango de estudio, es decir, tanto en el foco como a grandes distancias. La ecuación explícita de esta dependencia espacial es la siguiente:

$$\log RME_i = \rho + \log(1 + \alpha e^{-\beta {d_i}^2}) \tag{1.1}$$

En la Ecuación 1.1 el parámetro ρ está relacionado con la Razón de Mortalidad Estandarizada en la zona de referencia, es decir, a grandes distancias del foco. El riesgo relativo en el foco es α y β controla la rapidez del decaimiento exponencial.

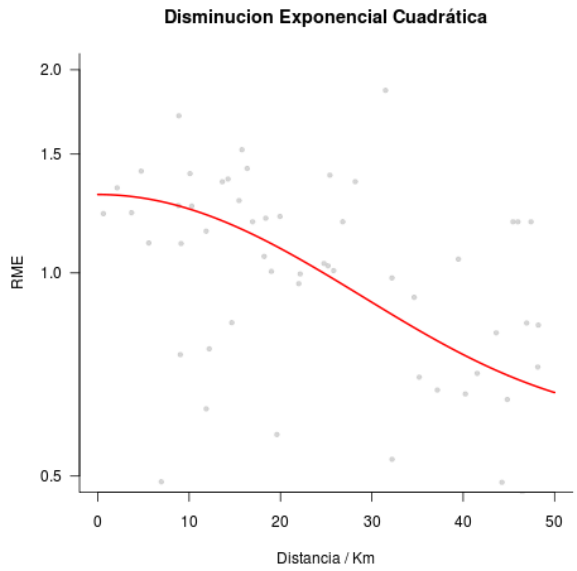


Figura 1.10: Modelo aditivo disminución exponencial cuadrática. Los puntos se corresponden a la RME para cada municipio y la linea roja representa la curva de riesgo

Otro modelo no lineal propuesto por Diggle et al. (1997) consiste en combinar el modelo anterior con un umbral en el alcance de la asociación. Se asume una zona expuesta en la que el riesgo es constante seguida de un decaimiento exponencial cuadrático hasta los niveles del riesgo en la zona de referencia (Figura 1.11). El decaimiento exponencial propuesto en estos dos últimos modelos contrasta con el salto discontinuo que presentan los modelos categóricos (Figuras 1.3 y 1.4), menos plausible desde el punto de vista epidemiológico.

La parametrización de este modelo (Ecuación 1.2) es muy similar al anterior y la interpretación de los parámetros es la misma. La principal diferencia es la introducción del parámetro δ, que marca la distancia a la

que el riesgo deja de ser constante (con valor e^{ρ}) para empezar a disminuir.

$$\log RME_i = \begin{cases} \rho & \text{si } d_i <= \delta \\ \rho + \log(1 + \alpha e^{-\beta {d_i}^2}) & \text{si } d_i > \delta \end{cases} \tag{1.2}$$

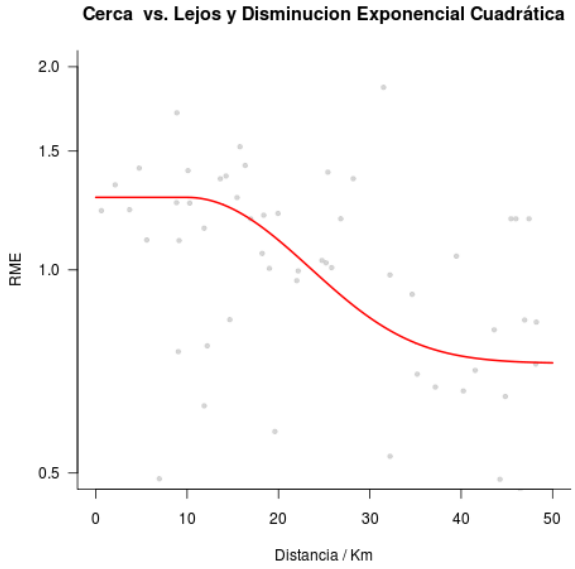


Figura 1.11: Modelo aditivo cerca vs. lejos con disminución exponencial cuadrática. Los puntos se corresponden a la RME para cada municipio y la linea roja representa la curva de riesgo

Como se ha comentado, el principal inconveniente de estas metodologías basadas en modelos no lineales radica en su implementación. Si bien la estimación de los parámetros puede llevarse a cabo mediante procedimientos de optimización estándar, evaluar su variabilidad requiere métodos más sofisticados (simulaciones de Monte Carlo) y, sobre todo, más costosos desde el punto de vista computacional (Seber and Wild 1989). Además, la comparación entre los distintos modelos no se puede realizar de una manera directa,

como sí ocurre en el caso de los modelos lineales. Este último aspecto resulta crucial a la hora de contrastar la existencia de las distintas componentes espaciales del riesgo.

Modelos no paramétricos Los métodos presentados hasta aquí consisten en diversas parametrizaciones de las variables espaciales en una regresión, cada una con sus características con respecto a la forma de la dependencia espacial del riesgo. También es posible hacer uso de las herramientas de suavización no paramétricas (modelos aditivos generalizados, splines, ...) para tratar estas variables (Bowman and Azzalini 1997; Hastie and Tibshirani 1990). Este tipo de enfoque es muy versátil a la hora de modelizar la relación del riesgo con la distancia. En general, no depende de decisiones arbitrarias, permite relaciones complejas (no lineales) y no presenta problemas a grandes distancias.

La principal complejidad de los modelos no paramétricos consiste en seleccionar el grado de suavización de la tendencia. Éste puede ir desde un modelo puramente lineal, que proporcionaría una suavización total, obteniendo una baja varianza a costa de un gran sesgo, a un modelo de interpolación que pase por cada punto de la nube, sin suavización, disminuyendo el sesgo a costa de aumentar la varianza. El criterio para seleccionar el grado de suavización suele ser minimizar el error cuadrático medio, obteniendo un buen balance entre sesgo y varianza. Las Figuras 1.12 y 1.13 presentan el resultado de estimar un modelo aditivo generalizado muy y poco suavizado respectivamente.

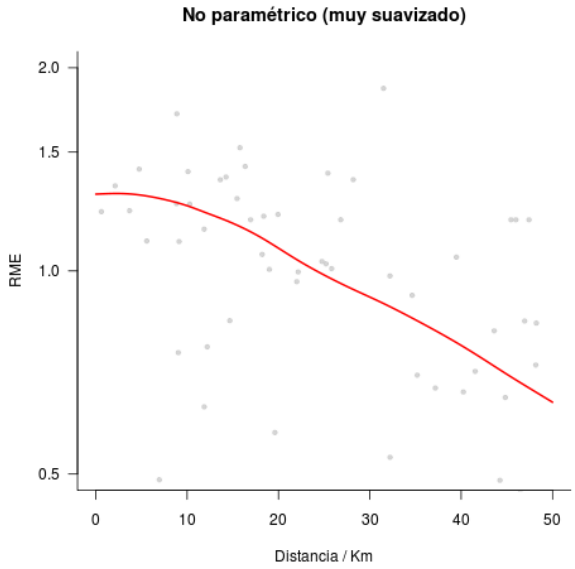


Figura 1.12: Modelos no paramétricos (muy suavizado) para la distancia. Los puntos se corresponden a la RME para cada municipio y la linea roja representa la curva de riesgo

La principal limitación de los métodos no-paramétricos reside en su interpretación: Si bien se puede inspeccionar visualmente el comportamiento espacial del riesgo, no se obtienen estimadores sobre los que realizar inferencia directa. Es decir, resulta difícil cuantificar y contrastar la existencia de una asociación espacial.

A destacar que los métodos no paramétricos permiten tener en cuenta la posición, como se verá a continuación.

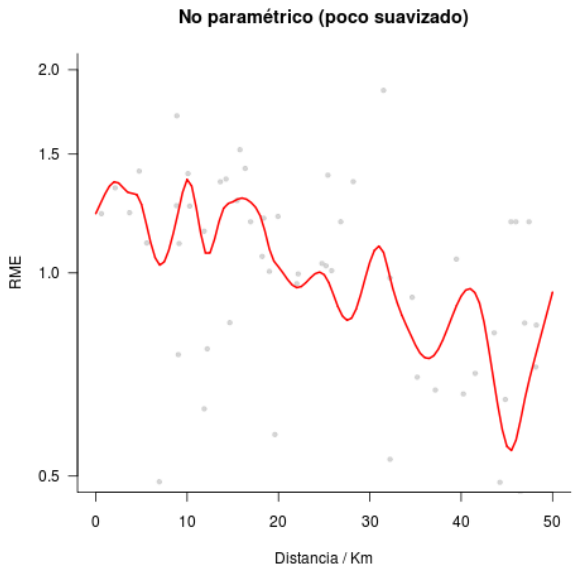


Figura 1.13: Modelos no paramétricos (poco suavizado) para la distancia. Los puntos se corresponden a la RME para cada municipio y la linea roja representa la curva de riesgo

Modelos de regresión basados en la localización geográfica

Hasta ahora los métodos presentados reducen la información acerca de la ubicación a la distancia con respecto al foco. Esta simplificación facilita los análisis, pero puede resultar insuficiente, ya que establece un riesgo que no varía con la dirección (isótropo). No es difícil imaginar situaciones en las que la distribución espacial del riesgo no sólo dependa de la distancia. La propia orografía (accidentes geográficos, cuencas fluviales, costa) y/o las condiciones meteorológicas predominantes (vientos) de cada escenario pueden condicionar la distribución espacial de manera anisótropa con respecto a la fuente contaminante. La distribución desigual del tamaño de los puntos

en la Figura 1.1 parece apuntar a que la relación del riesgo con la distancia no sea la misma para todas las direcciones.

La única referencia a modelos paramétricos que exploren la asociación con el ángulo se encuentra introducida en Lawson (1993) y aplicada en Biggeri et al. (1996). Se trata de introducir el seno y el coseno del ángulo en la regresión. De esta manera, se consigue modelizar en parte la componente angular del riesgo. Aún así, la parametrización que proponen no se ha desarrollado de manera coherente, teniendo en cuenta un modelo sencillo, plausible e interpretable.

Los métodos no paramétricos permiten modelizar simultáneamente la relación del riesgo con las dos coordenadas espaciales, latitud y longitud (Lawson et al. 1999). Así se puede obtener una descripción que tiene en cuenta la posición con respecto al foco, no sólo la distancia. Es una generalización de una curva de riesgo (presentada en los métodos anteriores) a una superficie de riesgo. El resultado es una figura como la Figura 1.14, muy útil para realizar una inspección visual pero con limitaciones a la hora de cuantificar la posible asociación entre la posición y la enfermedad.

En esta tesis se desarrollan modelos anisótropos que tienen en cuenta tanto la distancia como la direccionalidad, resultando en tendencias de riesgo muy flexibles y cuyos coeficientes son directamente interpretables.

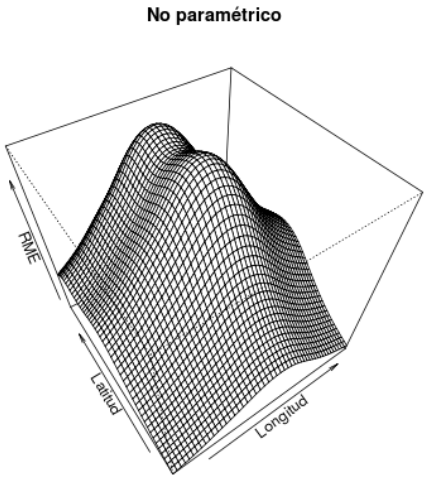


Figura 1.14: Métodos no paramétricos para la posición. Se presenta la superficie de riesgo, en la que la altura es proporcional a la RME

CAPÍTULO 2

Hipótesis y objetivos

2.1. Hipótesis

Los indicadores de enfermedad de las poblaciones residentes en entornos industriales presentan patrones espaciales con componentes radiales y direccionales asociados con la ubicación del foco contaminante.

El frecuente análisis exploratorio del impacto en salud de un foco contaminante suele basarse en una elección arbitraria de la distancia umbral. Esta distancia puede estimarse a partir de los propios datos.

La utilización de modelos isótropos supone una mala clasificación de la exposición a contaminantes industriales que impide poner en evidencia los excesos de riesgo. Los modelos anisótropos hacen un uso más eficiente y flexible de la localización geográfica como medida de exposición en este tipo de estudios ecológicos.

2.2. Objetivos

1. Desarrollar metodologías que permitan, ampliando las existentes, estimar y cuantificar la asociación espacial entre un presunto foco contaminante y la enfermedad en su entorno, teniendo en cuenta la localización geográfica de las poblaciones estudiadas. Una metodología que permita:

   a) Estimar el alcance de la asociación espacial a partir de los datos (componente radial).

   b) Establecer distribuciones espaciales direccionales (componente direccional).

   c) Contrastar la existencia de cada uno de las asociaciones por separado y de manera conjunta.

   d) Aplicarse de manera sistemática a grandes conjuntos de datos en estudios exploratorios.

2. Comprobar el rendimiento de las nuevas propuestas y compararlo con las existentes en situaciones controladas.

3. Aplicar la metodología al estudio de la mortalidad por diversos tipos de cáncer en el entorno de industrias contaminantes.

CAPÍTULO 3

Metodología

3.1. Consideraciones generales

Los métodos tradicionales presentan una serie de inconvenientes que muchas veces se aceptan como inevitables. En este capítulo se pretende dar a conocer nuevos métodos y parametrizaciones que permitan salvar algunas de esas limitaciones.

El principal obstáculo que presentan los métodos no paramétricos es su interpretación e inferencia sobre los resultados. A consecuencia de ello, en adelante, el abordaje se hará desde un punto de vista paramétrico, que permite cuantificar la relación entre la posición y el riesgo.

Como se ha visto, algunas de las técnicas utilizadas de manera sistemática tienen cierta dosis de arbitrariedad. Los métodos basados en la categorización de la distancia (el análisis cerca vs. lejos por ejemplo) necesitan una definición a priori de los puntos de corte. La variación de este criterio

puede condicionar drásticamente los resultados obtenidos. Por otra parte, simplificar la relación del riesgo con la distancia puede restar plausibilidad epidemiológica a los resultados. Una tendencia perfectamente lineal o un escalón no siempre reflejan de manera correcta la realidad. Aquí se proponen modelos más flexibles a la hora de modelizar la relación del riesgo con la distancia, en los que se evita la arbitrariedad de elegir una distancia de alcance a priori, estimándose ésta a partir de los datos.

En los entornos de focos contaminantes pueden darse asociaciones espaciales direccionales. Fenómenos atmosféricos y características orográficas son posibles ejemplos de factores que perturban la adireccionalidad del riesgo. Las aproximaciones más comunes no tienen en cuenta este tipo de situaciones, con lo que la estimación del riesgo asociado al foco contaminante puede verse sesgada. En esta sección se desarrollan extensiones de los modelos espaciales que tienen en cuenta tanto el factor radial (asociación con la distancia) como el direccional (asociación con la dirección).

Antes de presentar la metodología conviene enunciar las asunciones básicas y establecer una notación coherente que permita seguir con facilidad la evolución de los modelos propuestos. A continuación se define dicha notación y se describen las variables y parámetros comunes a todos ellos.

En adelante se asumirá un modelo de Poisson log-lineal en el que el número observado de casos O_i en el área i sigue una distribución de Poisson con media $E_i\lambda_i$, donde E_i es el número esperado de casos en el área i bajo las tasas específicas por edad de la población de referencia y λ_i es un factor multiplicativo que depende tanto de la posición relativa del área i respecto

al foco (X_i) como de otras covariables de ajuste $z_{ij}, j = 1, .., J$.

Las asunciones del modelo se presentan en la Ecuación 3.1 donde $f(X_i|\boldsymbol{\theta})$ es una función de dependencia espacial sobre X_i con parámetros $\boldsymbol{\theta}$ (incluido el intercept) y δ_j es el coeficiente asociado a la covariable z_{ij} que representa el logaritmo de la razón de tasas ajustadas (en adelante riesgo relativo) para cada incremento de una unidad de dicha covariable (Clayton and Hills 1993).

$$\begin{aligned} O_i &\sim Poisson(E_i\lambda_i) \\ log\lambda_i &= f(X_i|\boldsymbol{\theta}) + \sum_{j=1}^{J} \delta_j z_{ij} \end{aligned} \tag{3.1}$$

Se asumirá también, sin pérdida de generalidad, que el foco se encuentra en el origen de coordenadas $X = (x = 0, y = 0)$. Por conveniencia, la posición se parametrizará mediante coordenadas polares.

$$X_i = (d_i, a_i) = \left(\sqrt{{x_i}^2, {y_i}^2}, \ \text{arc tg} \frac{y_i}{x_i} \right) \tag{3.2}$$

En la ecuación 3.2 x_i e y_i son las coordenadas cartesianas (latitud y longitud, por ejemplo), d_i es la distancia en linea recta hasta el foco y a_i es el ángulo formado con la horizontal.

Las secciones que vienen a continuación sirven para motivar, presentar y describir la metodología propuesta. Para el desarrollo pormenorizado de la parametrización se remite al Apéndice A.

3.2. Punto de partida: Modelo lineal (radial)

El punto de partida de las innovaciones propuestas es el modelo lineal, que ya ha sido presentado en la Sección 1.4 del Capítulo 1. En él se asume un decrecimiento del riesgo proporcional la distancia. La Ecuación 3.3 presenta una reparametrización del modelo radial que, además de presentar una interpretación de los coeficientes más intuitiva, facilita la posterior generalización de la función de dependencia espacial.

$$f_{radial}(d_i|\beta_0, \beta_1) = \beta_0 + \beta_1 \left(1 - \frac{d_i}{d_{ref}}\right) \tag{3.3}$$

La interpretación de los coeficientes es sencilla:

- β_0 es el logaritmo de la RME en la distancia de referencia (d_{ref}). Se puede asumir que la distancia de referencia sea la máxima distancia al foco dentro de la zona de estudio ($d_{ref} = d_{max}$)
- β_1 es el logaritmo del riesgo relativo en el foco respecto a la distancia de referencia

El modelo es lineal en los parámetros, con lo que la estimación se reduce a un análisis de regresión lineal generalizada.

Si se unen mediante una linea todos los puntos que presentan un mismo riesgo se obtiene una circunferencia. Variando los valores del riesgo se van obteniendo circunferencias concéntricas de diversos radios, centradas en el foco. Estas *curvas de nivel de riesgo* o lineas *isorriesgo* facilitan la visualización e interpretación del modelo y serán utilizadas en adelante.

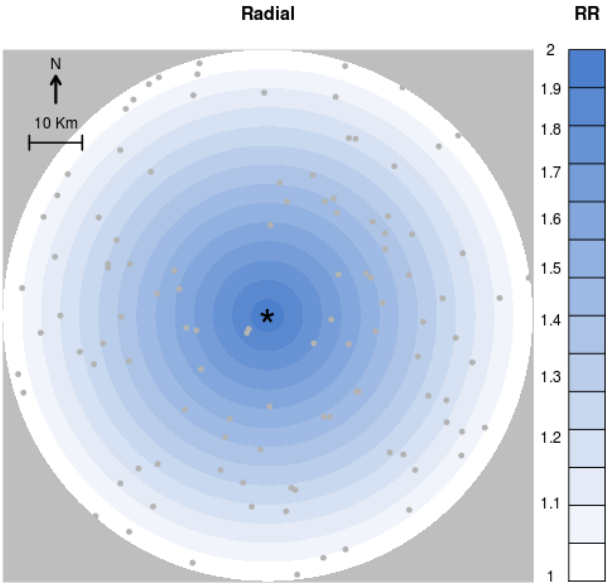


Figura 3.1: Superficie de riesgo al rededor de un foco estimada por el modelo radial. La escala de colores se corresponde con la intensidad del riesgo relativo. Los puntos indican la posición de las áreas de estudio (municipios) y el asterisco (*) la localización del foco. Se toma como referencia la RME a 50 Km (distancia máxima)

Los riesgo relativos (codificados mediante la escala de colores) de la Figura 3.1 decrecen de manera continua. De hecho, en las zonas mas alejadas este modelo estimaría un riesgo relativo de 0. Dado que la ausencia de riesgo se corresponde con el valor 1, esta modelización asigna a las zonas más alejadas ¡un efecto protector infinito! Más adelante se propone como solventar esta limitación.

Por otra parte, la Figura 3.1 presenta simetría radial: el riesgo estimado sólo depende de la distancia al foco, no de la dirección. Como ya se ha comentado, pueden existir situaciones en las que esta asunción no sea aceptable.

3.3. Modelo radial con umbral

Resulta interesante ampliar el modelo de partida de manera que permita una modelización más realista de la dependencia con la distancia sin renunciar a la simplicidad del mismo. Para ello, se propone introducir un término que restringa la asociación espacial espacial del foco, estableciendo un alcance máximo a partir del cual la relación espacial se anule. Éste alcance máximo define una zona de referencia en la que la componente espacial de la estimación de la RME es constante.

$$f_{radial+umbral}(d_i|\beta_0, \beta_1, \lambda) = \beta_0 + \beta_1 \left(1 - \frac{d_i}{\lambda}\right)^+ \tag{3.4}$$

con

$$x^+ = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } x >= 0 \end{cases} \tag{3.5}$$

En este caso, los coeficientes se interpretan de la siguiente manera:

- λ es el alcance de la asociación espacial del foco
- β_0 es el logaritmo de la RME más allá de la distancia definida por λ
- β_1 es el logaritmo del riesgo relativo en el foco respecto a la zona de referencia

La parametrización propuesta no es lineal. Esto presenta inconvenientes a la hora de ajustar el modelo (Seber and Wild 1989). El procedimiento de estimación de la verosimilitud y los coeficientes se detalla en el Apéndice A.

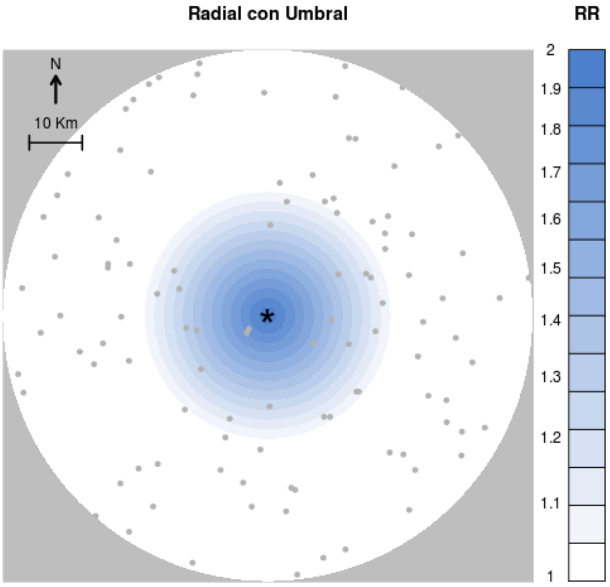


Figura 3.2: Superficie de riesgo al rededor de un foco estimada por el modelo radial con umbral. La escala de colores se corresponde con la intensidad del riesgo relativo. Los puntos indican la posición de las áreas de estudio (municipios) y el asterisco (*) la localización del foco

La descripción que proporciona este modelo es idéntica a la del modelo radial dentro del rango de alcance (λ). Fuera de él se anula la relación espacial. También tiene en cuenta la continuidad del riesgo; éste decrece hasta el alcance y para mayores distancias mantiene un valor constante.

Con esto se consigue una descripción más verosímil desde el punto de vista epidemiológico. Por otra parte, la Figura 3.2 continúa presentando simetría radial. No se están teniendo en cuenta posibles componentes direccionales en la asociación espacial entre el riesgo y el foco contaminante.

3.4. Modelo anisótropo

La siguiente mejora a introducir en la modelización radial es permitir que el riesgo presente direccionalidad con respecto al foco. Observando las figuras con lineas isoriesgo circulares, se plantea la posibilidad de generalizar esta situación y "deformar" los círculos para que dejen de ser simétricos. La deformación asimétrica de la circunferencia más sencilla es la elipse. De manera intuitiva, una elipse es una circunferencia que se ha estirado en una dirección determinada, reduciéndose en la perpendicular.

Para conseguir lineas isorriesgo elípticas es necesario hacer uso de una variable angular (a_i). La parametrización debe de ser capaz de describir la dirección del "alargamiento" y su magnitud (es decir, el grado de asimetría), que se conoce como excentricidad. En concreto, partiendo del modelo radial (Ecuación 3.3), se trata de introducir una dependencia angular en la distancia que define el alcance de la asociación (d_{ref}). Teniendo esto en cuenta se propone el siguiente modelo

$$\begin{aligned} f_{anisotropo}(d_i, a_i|\beta_0, \beta_1, \epsilon, \omega) = &\ \beta_0 + \beta_1 \left(1 - \frac{d_i}{u(a_i|\epsilon, \omega)}\right) \\ u(a_i|\epsilon, \omega) = &\ \frac{d_{ref}}{1 - \epsilon cos(a_i - \omega)} \end{aligned} \tag{3.6}$$

Los coeficientes del modelo tienen la siguiente interpretación

- β_0 es el logaritmo de la RME en la distancia de referencia (d_{ref}) en la dirección perpendicular a ω. Se puede asumir que la distancia de referencia sea la máxima distancia al foco dentro de la zona de estudio ($d_{ref} = d_{max}$)

- β_1 es el logaritmo del riesgo relativo en el foco respecto a la distancia de referencia
- ω es la dirección de riesgo máximo, aquella en la que decae de manera más gradual
- ϵ es la excentricidad de la elipse, relacionada directamente con el grado de asimetría del riesgo. Una circunferencia es una elipse con excentricidad nula

La propuesta esta inspirada y guarda similitud con Lawson (1993). La motivación, desarrollo e interpretación aquí presentada supone una extensión y una mejora de ésta.

Esta parametrización no es lineal, pero se puede reparametrizar de manera que sí lo sea (ver el Apéndice A), simplificando el procedimiento de estimación a un análisis de regresión lineal generalizada. Se ha presentado aquí la versión no lineal porque facilita la interpretación de los parámetros.

Mediante la variable angular a_i y un par de parámetros (ω y ϵ) que describen la dirección e intensidad de la asimetría, se consigue generalizar el modelo radial para permitir direccionalidad en el riesgo. Las lineas isorriesgo de la Figura 3.3 son elipses con el foco contaminante en uno de los focos geométricos y el otro en la dirección de ω.

Esta modelización más flexible permite describir escenarios en los que exista anisotropía en el riesgo. Además, cuantifica el grado de direccionalidad (mediante ϵ) y la dirección de máximo riesgo (vía ω). Sin embargo, hereda una de las limitaciones del modelo radial: el alcance de la asociación espacial

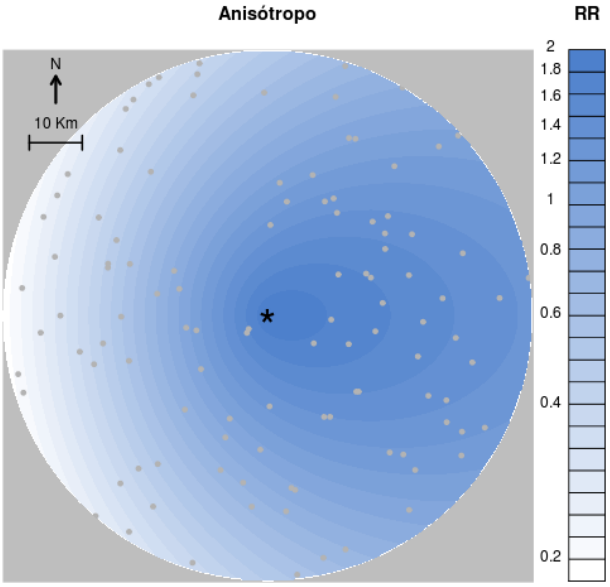


Figura 3.3: Superficie de riesgo al rededor de un foco estimada por el modelo anisótropo. La escala de colores se corresponde con la intensidad del riesgo relativo. Los puntos indican la posición de las áreas de estudio (municipios) y el asterisco (*) la localización del foco. Se toma como referencia la RME a 50 Km (distancia máxima)

no está limitado, con lo que el modelo pierde validez a grandes distancias y, con ello, plausibilidad epidemiológica.

3.5. Modelo anisótropo con umbral

Las ampliaciones del modelo radial propuestas hasta ahora atacaban el problema del alcance y la direccionalidad de manera independiente. El siguiente paso en la evolución de esta metodología consiste en combinar ambas propuestas. Con ello se podrán describir entornos de focos contaminantes en los que la asociación espacial no sea la misma en todas direcciones y además

presente un alcance limitado.

Se trata de describir la situación mediante elipses isorriesgo que, llegadas a un alcance que dependa del ángulo, converjan a la región de referencia.

$$\begin{aligned} f_{anisotropo+umbral}(d_i, a_i|\beta_0, \beta_1, \lambda, \epsilon, \omega) = \ & \beta_0 + \beta_1 \left(1 - \frac{d_i}{u(a_i|\epsilon, \omega)}\right)^+ \\ u(a_i|\lambda, \epsilon, \omega) = & \frac{\lambda}{1 - \epsilon cos(a_i - \omega)} \end{aligned} \tag{3.7}$$

en donde $(x)^+$ se define igual que en la ecuación 3.5 y los coeficientes heredan la interpretación de los modelos precedentes:

- λ es el alcance de la asociación espacial del foco en la dirección perpendicular a ω
- β_0 es el logaritmo de la RME en la distancia de referencia
- β_1 es el logaritmo del riesgo relativo en el foco respecto a la distancia de referencia
- ω es la dirección de riesgo máximo, aquella en la que decae de manera más gradual
- ϵ es la excentricidad de la elipse, relacionada directamente con el grado de asimetría del riesgo. Una circunferencia es una elipse con excentricidad nula

Este modelo establece como zona de referencia toda el área que se encuentra más allá de la elipse máxima. Es la generalización del modelo radial con umbral, pero en este caso la distancia de alcance depende del ángulo.

Como en el caso del modelo radial con umbral, la parametrización no es lineal. Para un análisis detallado de los procedimientos empleados en la estimación véase el Apéndice A.

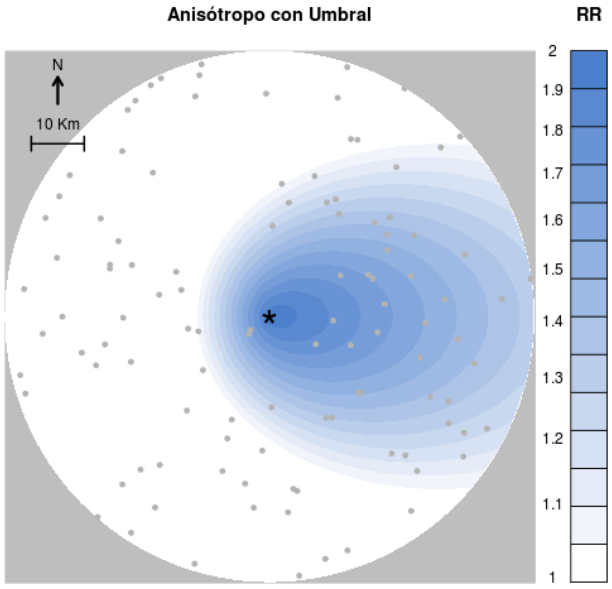


Figura 3.4: Superficie de riesgo al rededor de un foco estimada por el modelo anisótropo con umbral. La escala de colores se corresponde con la intensidad del riesgo relativo. Los puntos indican la posición de las áreas de estudio (municipios) y el asterisco (*) la localización del foco

Este modelo permite una descripción más verosímil desde el punto de vista epidemiológico, ya que incluye la posibilidad de que la asociación espacial del foco contaminante en su entorno dependa de la dirección y esté limitada dentro de un alcance. Además, condiciona la asociación espacial angular a la existencia de la radial. Esto resulta muy conveniente desde el punto de vista causal; sería difícil sostener causalidad en una situación en la que un foco presentara relación con la dirección pero no con la distancia.

3.6. Comparación de modelos

Es tan importante presentar y describir cada uno de los modelos como destacar la relación que existe entre ellos. La parametrización propuesta es anidada; a medida que los modelos son más complejos, sus predecesores se encuentran contenidos en ellos como casos particulares. Así, el modelo radial se corresponde con el radial con umbral si se fija el valor de λ en la distancia de referencia (d_{ref}) del escenario estudiado. También se recupera el modelo radial a partir del anisótropo fijando el valor de la excentricidad (ϵ) a 0. De la misma manera ($\epsilon = 0$), el modelo radial con umbral se puede recobrar a partir del anisótropo con umbral. Para reducir éste último al anisótropo, basta con igualar λ a la distancia de referencia. Por supuesto, si $\epsilon = 0$ y $\lambda = d_{ref}$ el modelo más complejo se reduce al radial.

Todo este galimatías es crucial a la hora de comparar modelos. Al estar anidados, se puede realizar directamente un contraste basado en la verosimilitud (Clayton and Hills 1993) para evaluar el rendimiento de cada uno (los detalles de implementación del contraste se presentan en el Apéndice A). Esto es de gran utilidad ya que permite contrastar la existencia de las distintas asociaciones espaciales y, con ello, crear un protocolo de decisión que determine que tipo de asociación se ajusta mejor a los datos. En concreto:

- Si alguno de los modelos propuestos resulta mejor que el nulo, existe una asociación espacial
- Si el modelo radial con umbral resulta mejor que el radial, la asociación espacial tiene un rango acotado

- Si el modelo anisótropo resulta mejor que el radial, existe, además, una asociación direccional
- Si el modelo anisótropo con umbral es mejor que los modelos radial con umbral y anisótropo, existe una asociación direccional con un rango acotado

El modelo nulo es aquél en el que no se incluye ninguna función de dependencia espacial (solo las covariables).

CAPÍTULO 4

Resultados

Los modelos presentados se han puesto a prueba en situaciones controladas para conocer y comparar su rendimiento (Apéndice B). No obstante, el interés final de las propuestas es su aplicación a escenarios reales, de manera que permitan ampliar y profundizar el estudio de la enfermedad en los entornos de focos contaminantes. En este capítulo se desarrolla la puesta en práctica, en situaciones de interés real, de la metodología introducida.

En lo sucesivo se considera el entorno como el circulo de 50 Km de radio alrededor del foco contaminante.

Las características de esta metodología permiten su aplicación de manera sistemática a grandes conjuntos de datos. Para demostrarlo, en esta sección se escogen las cinco causas tumorales con mortalidades más altas, en España, para hombres y mujeres por separado. También se presentan los resultados pormenorizados de una serie escogida de localizaciones específicas.

4.1. Material

El indicador de enfermedad que se estudia es la mortalidad a nivel municipal. Las localizaciones tumorales específicas más frecuentes para los hombres son pulmón, colon y recto, próstata, estómago y vejiga. Para las mujeres mama, colon y recto, estómago, pulmón y páncreas. Estos tipos de tumores supusieron el 95 % de las defunciones por cáncer en hombres y el 50 % en las mujeres en España durante el 2008 (Área de Epidemiología Ambiental y cáncer Centro Nacional de Epidemiología ISCIII 2008). Además, en la literatura se han descrito asociaciones con las exposiciones a contaminantes para la mayoría de ellos (ver Capítulo 1). El periodo de estudio comprende los años entre 1996 a 2005. Para cada municipalidad se obtuvieron los casos observados del INE y se calcularon los esperados a partir de las poblaciones municipales y las tasas nacionales, estratificadas por edad. Debido al importancia de las variables socioeconómicas como confusoras en el estudio de estas causas (Elliott et al. 1996), se incluyen en el análisis la tasas de analfabetismo y desempleo.

Las ubicaciones de los focos contaminantes se obtuvieron del registro de actividades industriales E-PRTR (*European Pollutant Release and Transfer Register*), creado por el Reglamento CE 166/2006, en España para el año 2007. Éste consiste en una relación de las emisiones a la atmósfera y al agua de todos los complejos industriales en los que se lleven a cabo una o más actividades que figuran en el anexo I de la Directiva 96/61/CE. Las actividades industriales registradas pertenecen a 6 grandes categorías:

1. Instalaciones de Combustión
2. Producción y transformación de metales
3. Industrias minerales
4. Industrias químicas
5. Gestión de residuos
6. Otras actividades (papeleras, tinte de textiles, cuero, mataderos, cría intensiva de aves y cerdos, instalaciones que utilizan disolventes orgánicos, fabricación de carbono o grafito)

Aquellos entornos en los que existe más de un foco de la misma categoría industrial y que comparten más del 50 % de la población se excluyen del análisis, para evitar parte de la posible distorsión debida a la presencia de focos múltiples. El análisis presentado se corresponde con las industrias que emiten al aire (447 localizaciones).

4.2. Aplicación sistemática

El procedimiento seguido es el siguiente: se ajustaron los modelos espaciales (*nulo, radial, radial con umbral, anisótropo y anisótropo con umbral*) para cada localización, tumor y sexo, comparándose los modelos obtenidos. La proporción de escenarios en los que se detecta alguna asociación espacial (significativa al 95 %) se puede consultar en la Tabla 4.1.

La columna denominada "No-Radial" indica la proporción de escenarios en los que no existe una asociación radial simple, pero sí asociaciones espa-

	Hombres				**Mujeres**		
	Espacial	Radial	No-Radial		Espacial	Radial	No-Radial
Pulmón	0.69	0.37	0.32	*Mama*	0.11	0.03	0.08
Colorrectal	0.31	0.15	0.16	*Colorrectal*	0.18	0.08	0.09
Próstata	0.09	0.03	0.06	*Estómago*	0.18	0.05	0.13
Estómago	0.19	0.07	0.12	*Pulmón*	0.16	0.10	0.06
Vejiga	0.19	0.08	0.11	*Páncreas*	0.06	0.02	0.04

Tabla 4.1: Proporción de entornos (sobre los 447 estudiados) que presentan una asociación espacial significativa al 95 %

ciales más complejas; es decir, situaciones en las que se detecta un umbral en el alcance de la asociación, direccionalidad en el riesgo o ambos fenómenos a la vez. Este valor es siempre similar al porcentaje de escenarios con relación radial detectados. Por tanto, utilizar los modelos propuestos duplica la tasa de detección de asociaciones espaciales del riesgo.

Dado que en este procedimiento se están llevando a cabo muchas pruebas de existencia a través de contrastes de hipótesis, los valores P asociados se han corregido controlando la tasa de falsos positivos (Benjamini 2001), de manera que éstos no invaliden los resultados.

4.3. Casos particulares

Se presenta aquí una muestra de casos elegidos entre aquellos, de los referidos en la sección anterior, en los que el modelo que mejor describe la situación es el *anisótropo con umbral*. La elección de los ejemplos mostrados viene en parte motivada por resultados anteriores que evidenciaron la asociación entre la mortalidad por distintos tipos de cáncer y la proximidad a complejos industriales. En concreto, tumores del aparato digestivo

(García-Pérez et al. 2010) y de vejiga (García-Pérez et al. 2009). Se pretende presentar una muestra que abarque los distintos tumores, localizaciones geográficas y categorías industriales. También se han elegido teniendo en cuenta las diferencias existentes entre sexos.

No se presentan resultados particulares para el cáncer de pulmón en hombres, ya que no se dispone de información a cerca del principal factor de confusión relacionado con esta causa: el consumo de tabaco.

Para descartar que la presencia de grandes núcleos poblacionales esté distorsionando los resultados, se realizó un análisis de sensibilidad excluyéndolos.

En todas las figuras de esta sección el diámetro de los puntos es proporcional al tamaño municipal, la intensidad del rojo a la RME cruda, las curvas de nivel azules son el resultado de ajustar el modelo anisótropo con umbral y los asteriscos verdes corresponden a las localizaciones de focos contaminantes.

Para cada localización se incluyen dos respresentaciones gráficas. La primera contiene, además de la distribución de los municipios, los focos contaminantes y el gradiente de riesgo codificado en una escala de color. La información que aporta la segunda es complementaria: aunque las curvas de nivel de riesgo son más difíciles de interpretar (sin el gradiente de colores) presenta una cartografía de la zona con el nombre de los municipios, la orografía y las principales vías de comunicación.

Para empezar con los casos particulares se estudia la mortalidad por

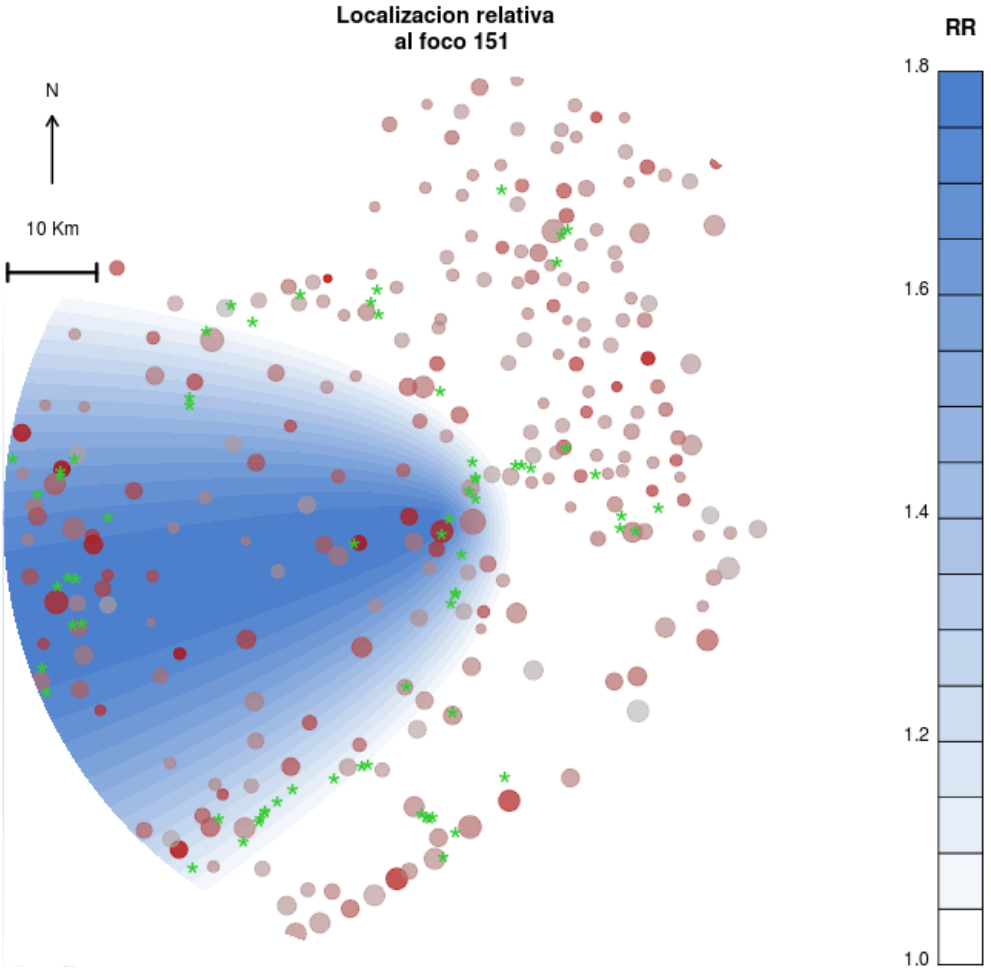
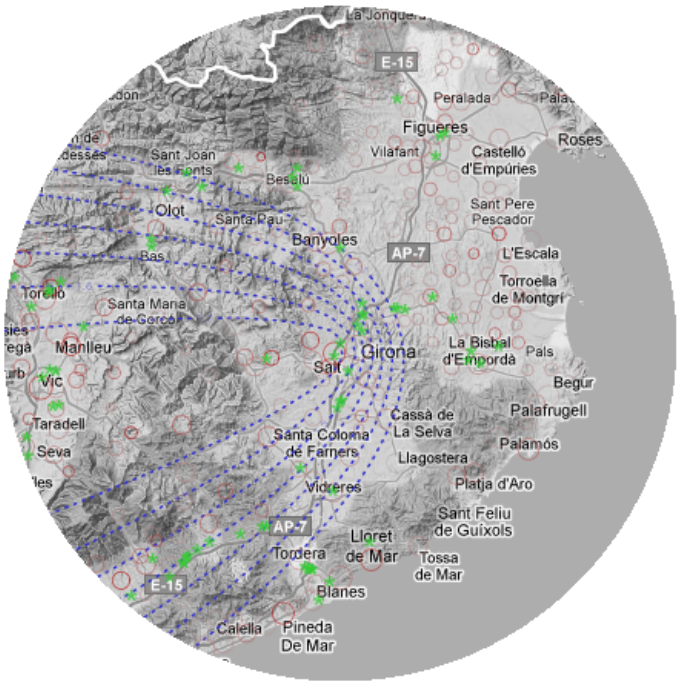


Figura 4.1: Riesgo de morir por cáncer de estómago en el entorno de una industria de Girona (hombres). Gradiente de riesgo

cáncer de estómago en el entorno de Girona. Dentro del municipio existen dos posibles focos contaminantes con actividades diferentes. Uno pertenece a la categoría de gestión de residuos y el otro es una fábrica de Nestlé catalogada dentro de las actividades alimentarias. El patrón de riesgo estimado por el modelo aplicado a cada foco es indistinguible.

Los resultados obtenidos en hombres (Figuras 4.1 y 4.2) y en mujeres



RRfoco 1.8 (1.43, 2.24)
alcance / Km 15 (10, 19)
excentricidad 1 (0.87, 1)
direccion / º 194 (167, 217)

Figura 4.2: Riesgo de morir por cáncer de estómago en el entorno de una industria de Girona (hombres). Mapa topográfico

(Figuras 4.3 y 4.4) son compatibles entre si. El riesgo relativo en el foco es el doble que en la zona de referencia (Tabla 4.2) y el patrón espacial de riesgo presenta direccionalidad acusada hacia el oeste, que podría estar relacionada con el régimen de vientos.

El patrón de mortalidad por cáncer de vejiga en hombres en el entorno de una fábrica de artículos pirotécnicos en Ronda (Málaga) presenta una

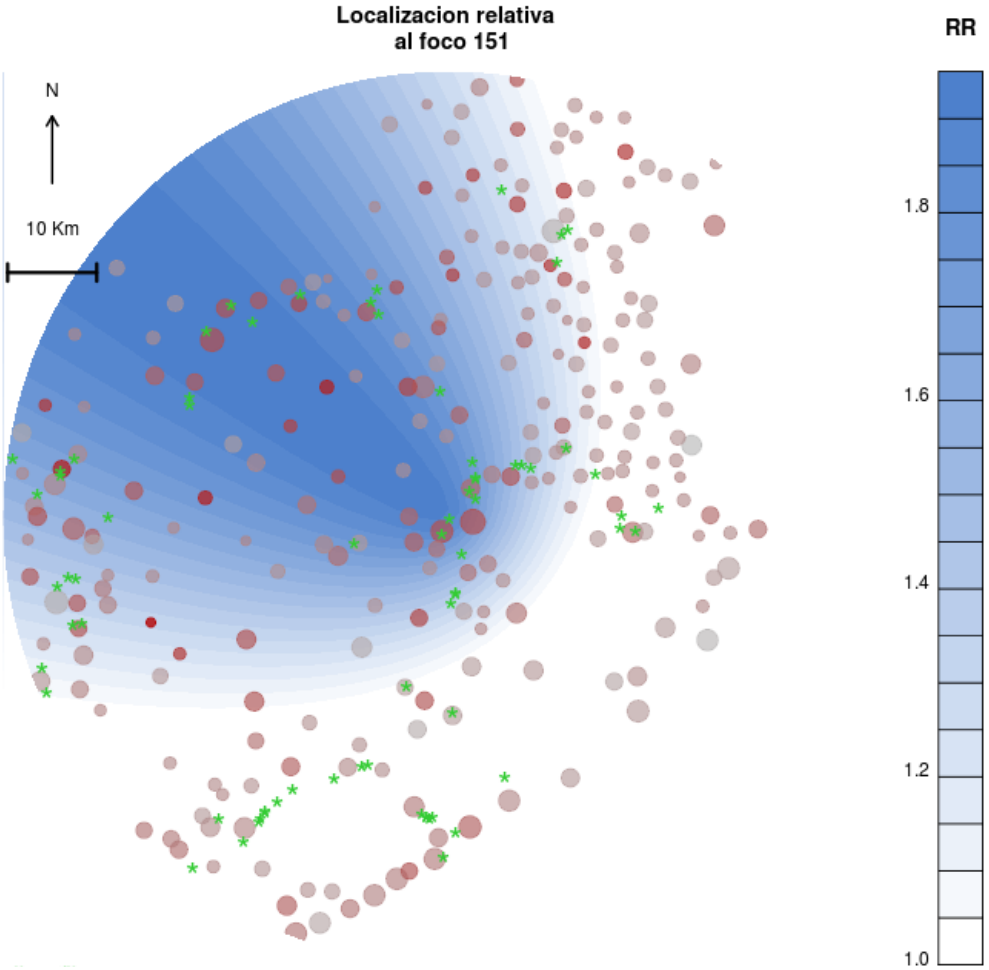
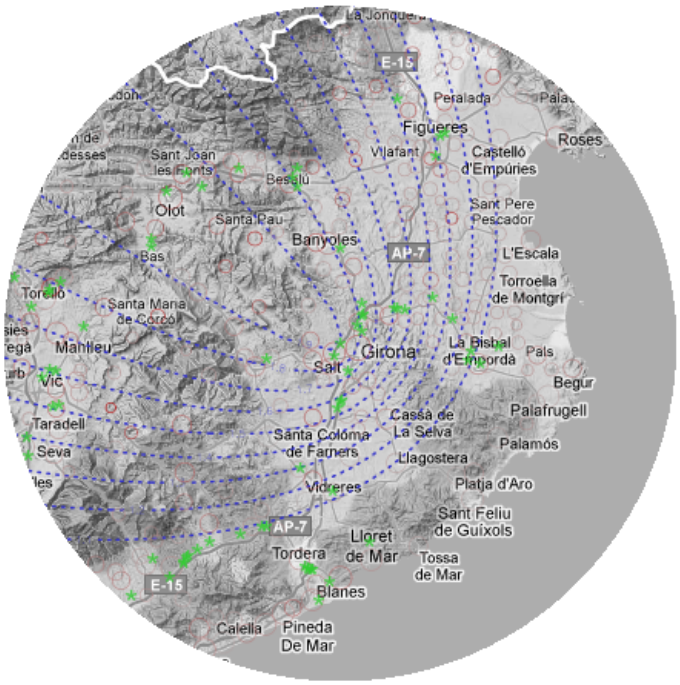


Figura 4.3: Riesgo de morir por cáncer de estómago en el entorno de una industria de Girona (mujeres). Gradiente de riesgo

Parámetro	**Hombres (IC 95 %)**	**Mujeres (IC 95 %)**
RR_{foco}	1.8 (1.43-2.24)	1.94 (1.5-2.44)
Alcance /Km	15 (10-19)	29 (20-36)
Excentricidad	1 (0.87-1)	1 (0.62-1)
Dirección / ^{o}	194 (167-217)	141 (120-176)

Tabla 4.2: Parámetros del modelo anisotropo con umbral para un entorno de Girona. Estimaciones puntuales e intervalo de confianza al 95 %



RRfoco 1.94 (1.5, 2.44)
alcance / Km 29 (20, 36)
excentricidad 1 (0.62, 1)
direccion / º 141 (120, 176)

Figura 4.4: Riesgo de morir por cáncer de estómago en el entorno de una industria de Girona (mujeres). Mapa topográfico

direccionalidad acusada hacia el norte (Figura 4.5). Este foco pertenece a la categoría de industria química. Es posible que la sierra de Ronda, situada al sur de municipio, tenga algo que ver con esta distribución (Figura 4.6). Las estimaciones de los parámetros se pueden consultar en la Tabla 4.3.

Siguiendo con la misma causa tumoral, estudiando el entorno de una papelera en Miranda de Ebro, se detecta una asociación con componente

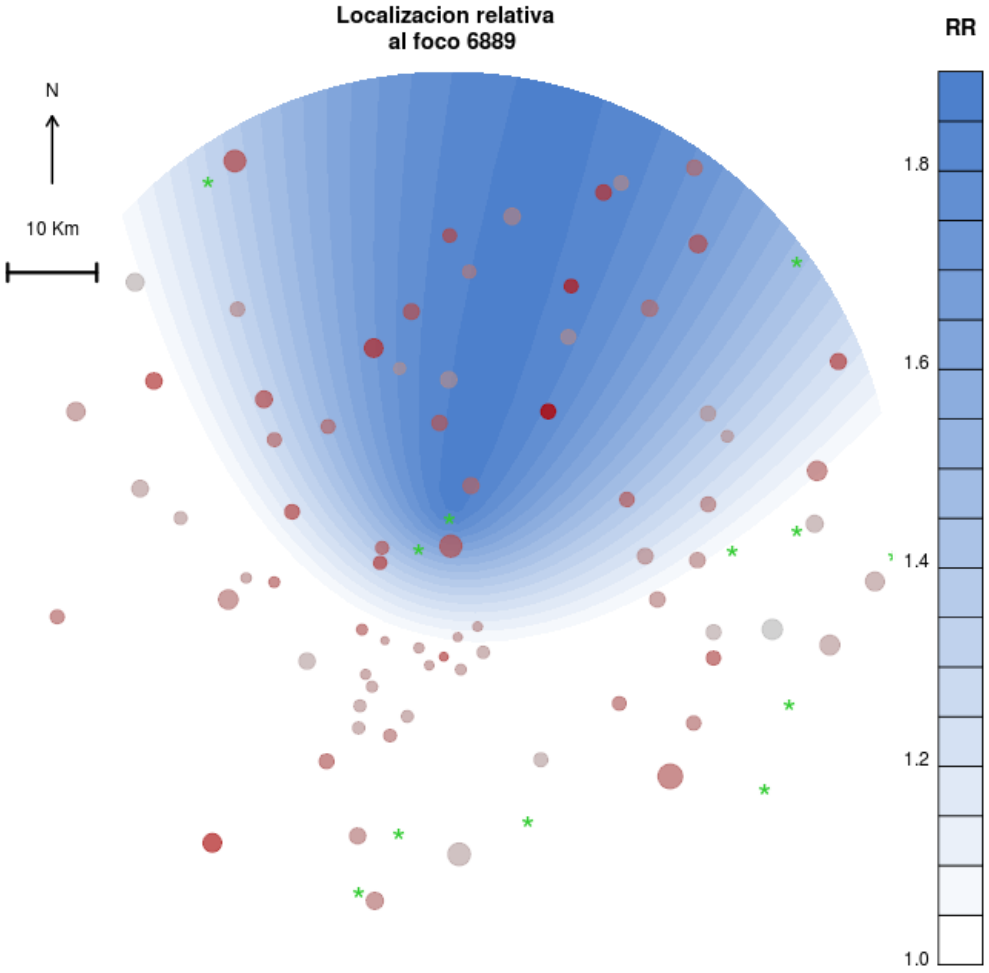


Figura 4.5: Riesgo de morir por cáncer de vejiga en el entorno de una industria de Ronda (hombres). Gradiente de riesgo

Parámetro	**Estimación**	**Lim. Inf. 95 %**	**Lim. Sup. 95 %**
RR_{foco}	1.87	1.21	2.73
Alcance /Km	29	14	41
Excentricidad	0.99	0.26	1.00
Dirección / ^{o}	75	47	135

Tabla 4.3: Parámetros del modelo anisotropo con umbral para un entorno de Ronda (hombres). Estimaciones puntuales e intervalo de confianza al 95 %

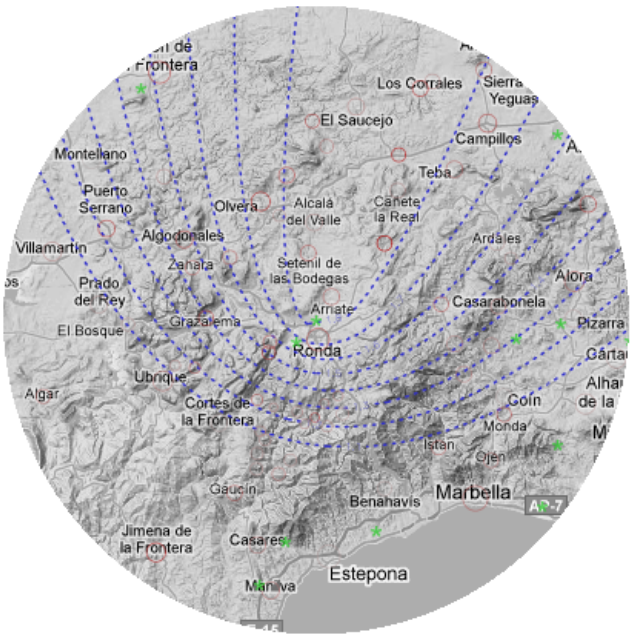


Figura 4.6: Riesgo de morir por cáncer de vejiga en el entorno de una industria de Ronda (hombres). Mapa topográfico

direccional como se puede ver en las Figuras 4.7 y 4.8. El exceso de riesgo detectado apunta al discurrir de las aguas Ebro abajo. Los parámetros del modelo ajustado están expuestos en la Tabla 4.4.

El siguiente escenario estudiado corresponde con el cáncer de colon y recto en mujeres en el entorno de la central térmica de As Pontes de García Rodríguez (A Coruña), que es una instalación de combustión. La estimación

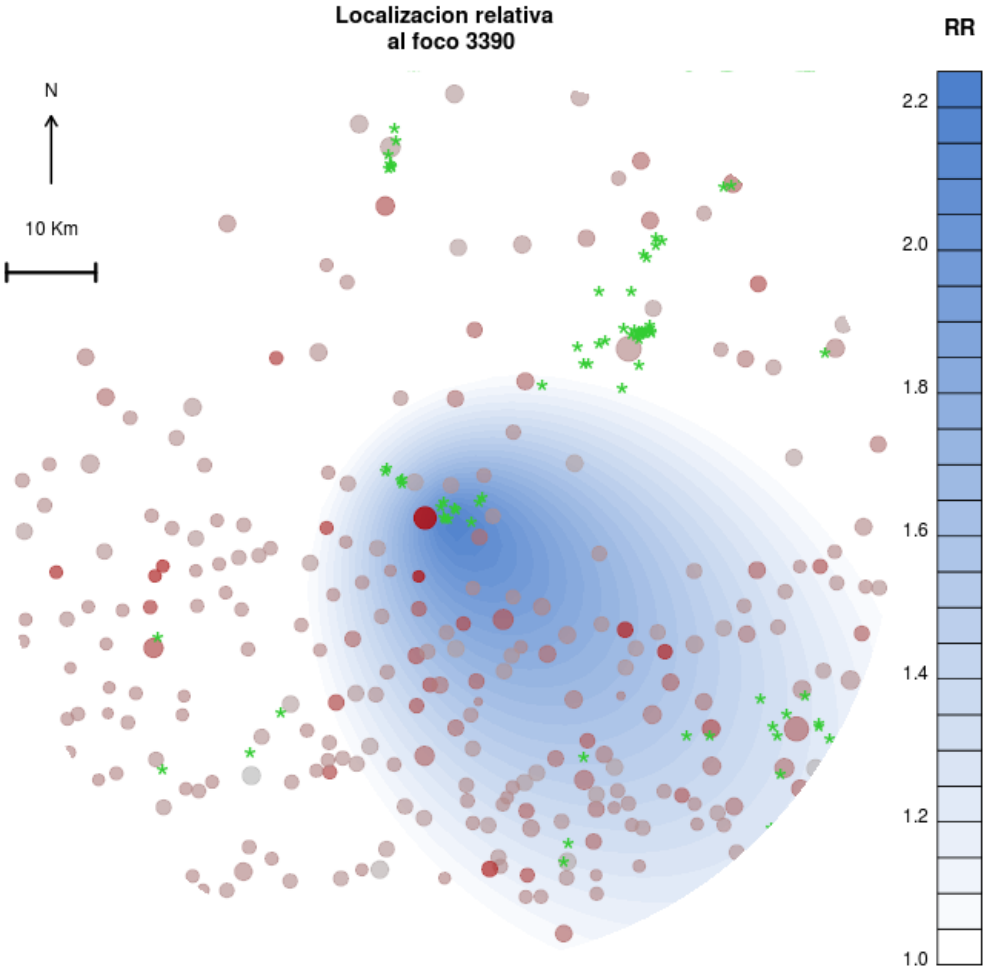


Figura 4.7: Riesgo de morir por cáncer de vejiga en el entorno de una industria de Miranda de Ebro (hombres). Gradiente de riesgo

Parámetro	**Estimación**	**Lim. Inf. 95 %**	**Lim. Sup. 95 %**
RR_{foco}	2.24	1.62	2.96
Alcance /Km	23	19	27
Excentricidad	0.67	0.53	0.69
Dirección / ^{o}	-44	-75	-4

Tabla 4.4: Parámetros del modelo anisotropo con umbral para un entorno de Miranda de Ebro (hombres). Estimaciones puntuales e intervalo de confianza al 95 %

Figura 4.8: Riesgo de morir por cáncer de vejiga en el entorno de una industria de Miranda de Ebro (hombres). Mapa topográfico

proporcionada por el modelo (Tabla 4.5) indica que el riesgo evoluciona hacia sureste. Es decir, sigue la extensión plana de la comarca Terra Chá hacia Villalba, contenido por el noroeste por los montes litorales (Figuras 4.9 y 4.10).

El patrón que se obtiene al estudiar el entorno de una industria metalúrgica de A Pontenova situada 70 Km al Este de As Pontes concuerda con

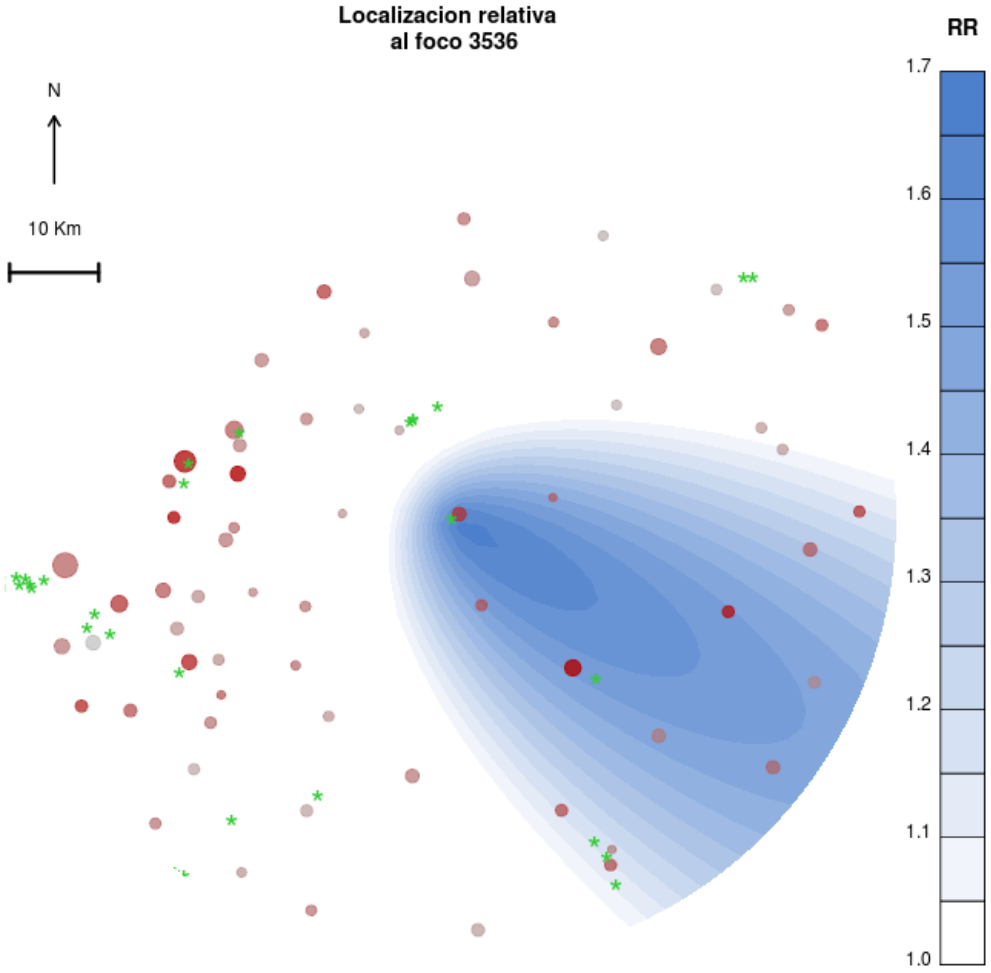
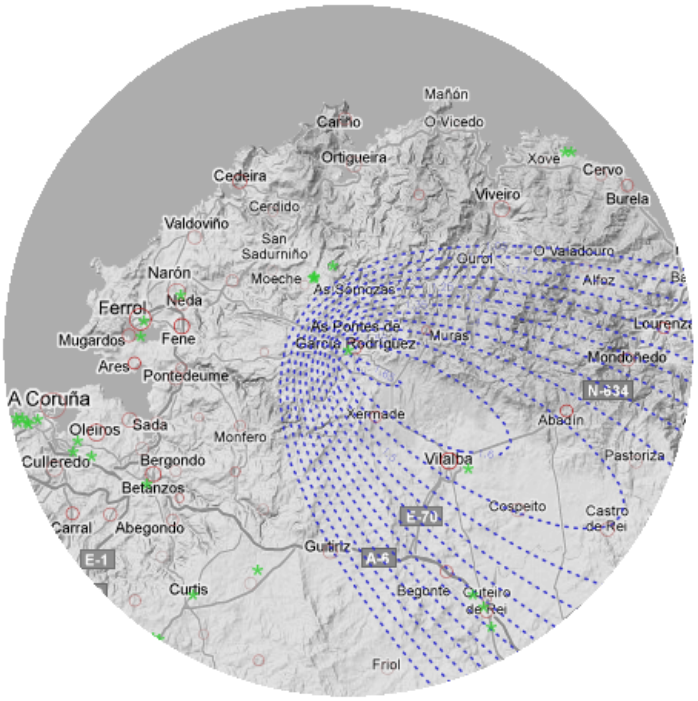


Figura 4.9: Riesgo de morir por cáncer colorrectal en el entorno de una industria de As Pontes de García Rodríguez (mujeres). Gradiente de riesgo

Parámetro	**Estimación**	**Lim. Inf. 95 %**	**Lim. Sup. 95 %**
RR_{foco}	1.68	0.99	2.65
Alcance /Km	13	0	21
Excentricidad	0.94	0.00	1.00
Dirección / ^{o}	-31	-111	21

Tabla 4.5: Parámetros del modelo anisotropo con umbral para un entorno de As Pontes de García Rodríguez (mujeres). Estimaciones puntuales e intervalo de confianza al 95 %



RRfoco 1.68 (0.99, 2.65)
alcance / Km 13 (0, 21)
excentricidad 0.94 (0, 1)
direccion / º -31 (-111, 21)

Figura 4.10: Riesgo de morir por cáncer colorrectal en el entorno de una industria de As Pontes de García Rodríguez (mujeres). Mapa topográfico

el anterior (Figuras 4.11 y 4.12 y Tabla 4.6) indicando que existe un área de exceso de riesgo en la planicie de la comarca Terra Chá que podría estar relacionada con cualquiera de los dos focos.

Se detecta una distribución direccional en el riesgo de morir por cáncer de pulmón en mujeres asociada con la cantera de Manilva (Málaga) (Figura 4.14). Es posible que el mar y la cadena montañosa que discurre paralela a

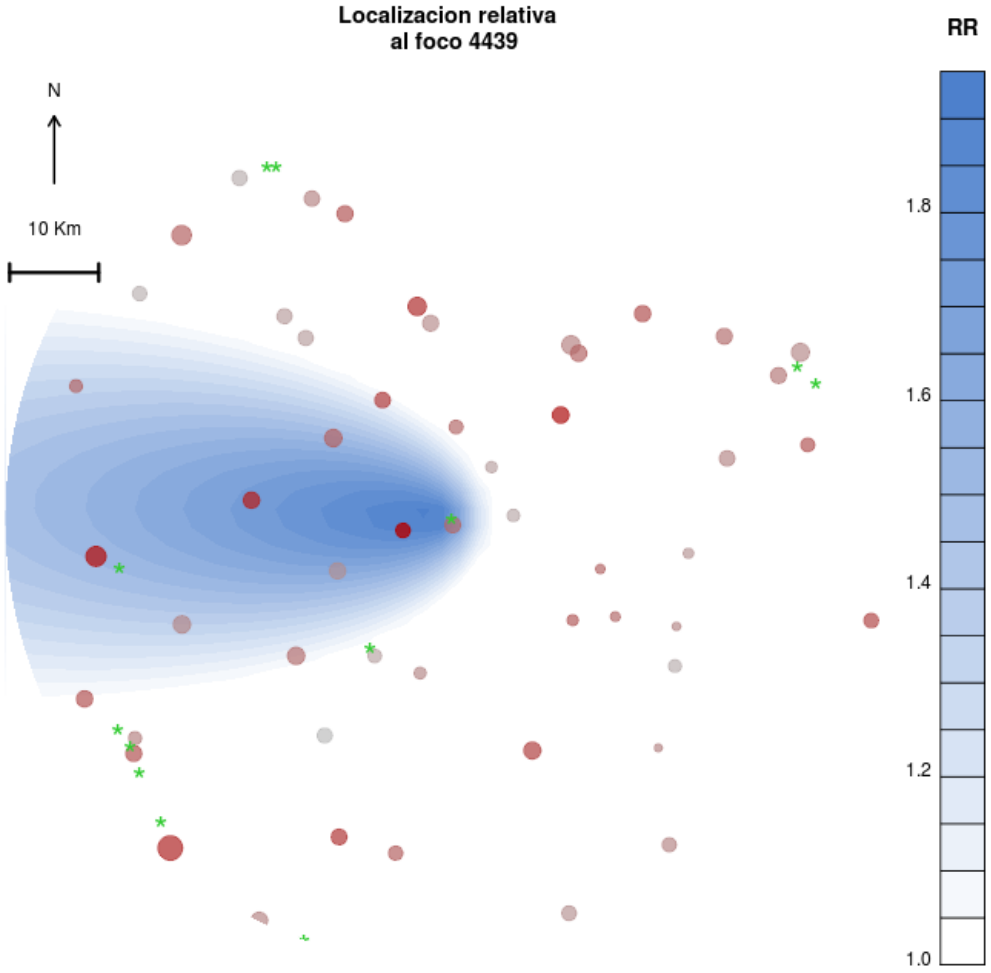


Figura 4.11: Riesgo de morir por cáncer colorrectal en el entorno de una industria de A Pontenova (mujeres). Gradiente de riesgo

Parámetro	**Estimación**	**Lim. Inf. 95 %**	**Lim. Sup. 95 %**
RR_{foco}	1.96	1.01	3.47
Alcance /Km	9	1	14
Excentricidad	0.92	0.05	1.00
Dirección / ^{o}	178	143	217

Tabla 4.6: Parámetros del modelo anisotropo con umbral para un entorno de A Pontenova (mujeres). Estimaciones puntuales e intervalo de confianza al 95 %

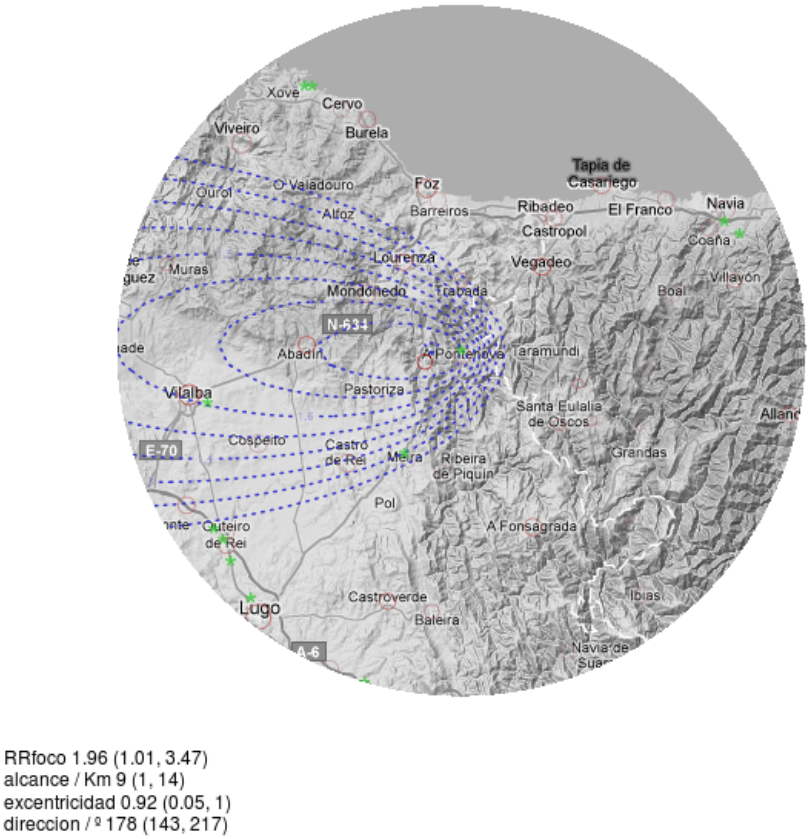


Figura 4.12: Riesgo de morir por cáncer colorrectal en el entorno de una industria de A Pontenova (mujeres). Mapa topográfico

la costa (Figura 4.14) sean responsables en parte de esta distribución.

Parámetro	**Estimación**	**Lim. Inf. 95 %**	**Lim. Sup. 95 %**
RR_{foco}	1.66	1.09	2.58
Alcance /Km	31	11	47
Excentricidad	0.83	0.19	1.00
Dirección / ^{o}	1	-53	81

Tabla 4.7: Parámetros del modelo anisotropo con umbral para un entorno de Manilva (mujeres). Estimaciones puntuales e intervalo de confianza al 95 %

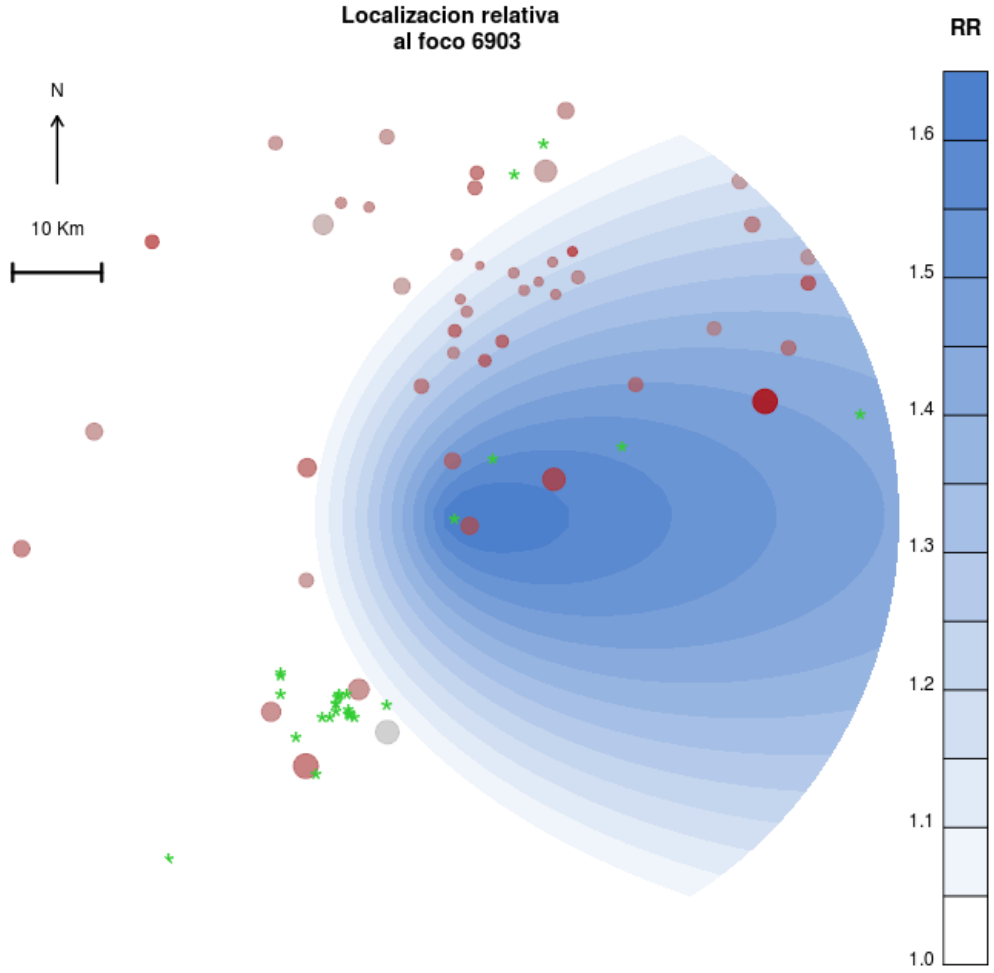


Figura 4.13: Riesgo de morir por cáncer de pulmón en el entorno de una industria de Manilva (mujeres). Gradiente de riesgo

La mortalidad por cáncer de próstata en los alrededores de Zamora presenta una asociación espacial con el vertedero de la ciudad (Figura 4.16). En la Figura 4.16 se aprecia que la dirección de máximo riesgo coincide con el cauce del río. También el viento, en el que predomina la componente este en la zona, puede estar contribuyendo. Los parámetros estimados se pueden consultar en al Tabla 4.8.

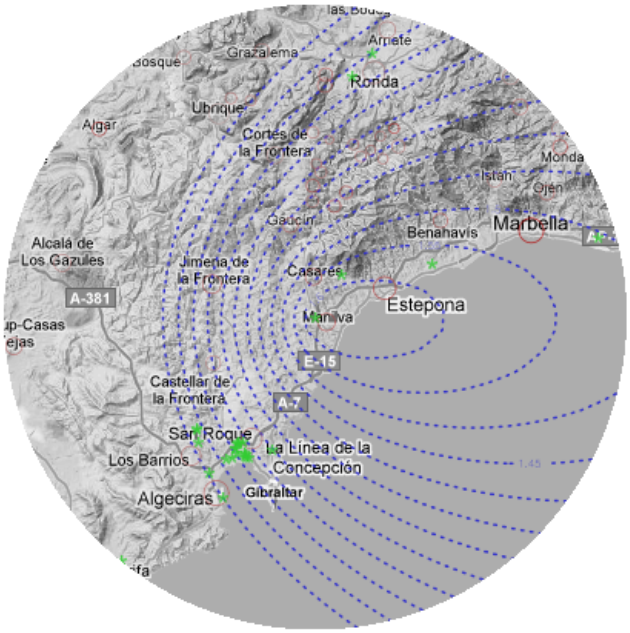


Figura 4.14: Riesgo de morir por cáncer de pulmón en el entorno de una industria de Manilva (mujeres). Mapa topográfico

Parámetro	**Estimación**	**Lim. Inf. 95 %**	**Lim. Sup. 95 %**
RR_{foco}	1.6	1.22	2.11
Alcance /Km	23	13	32
Excentricidad	0.77	0.13	1.00
Dirección / ^{o}	25	-81	75

Tabla 4.8: Parámetros del modelo anisotropo con umbral para un entorno de Zamora (hombres). Estimaciones puntuales e intervalo de confianza al 95 %

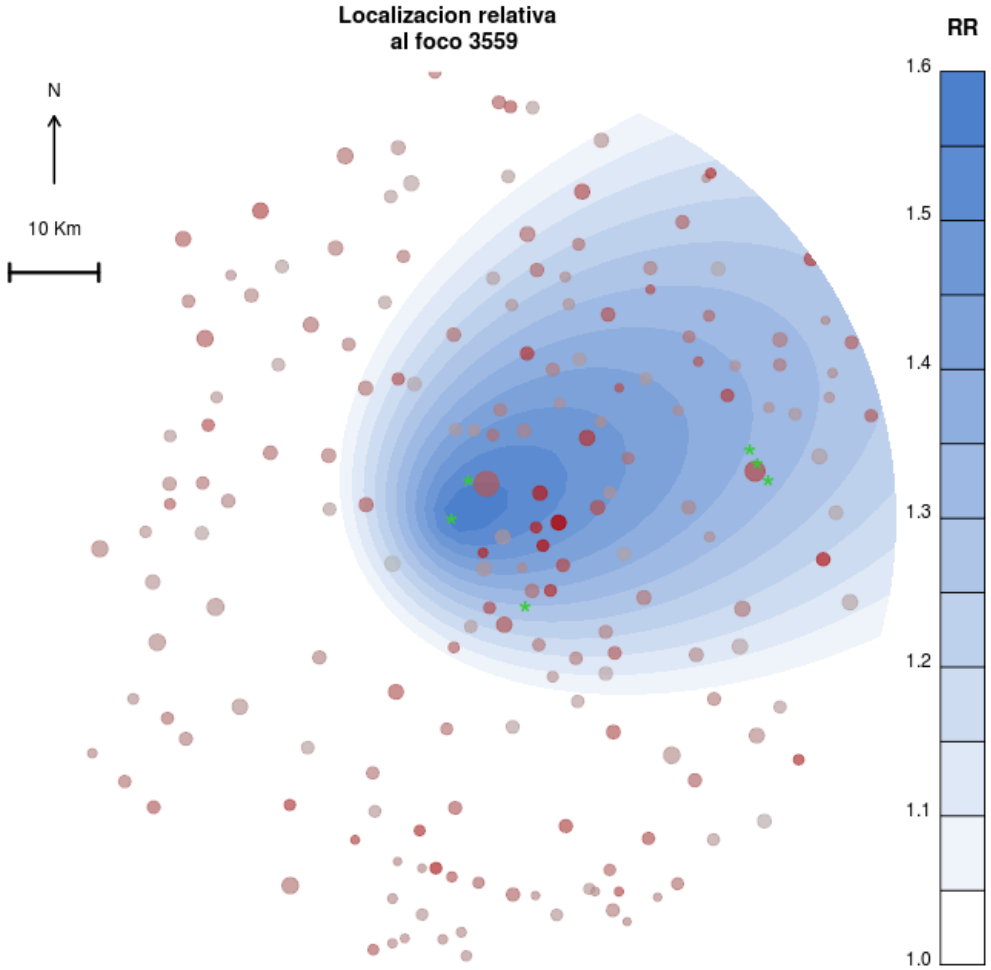


Figura 4.15: Riesgo de morir por cáncer de próstata en el entorno de una industria de Zamora (hombres). Gradiente de riesgo

La mina a cielo abierto de Santa Marta en Belorado presenta una asociación espacial con la mortalidad por cáncer de páncreas en las mujeres de su entorno (Tabla 4.9, Figura 4.17). Como se aprecia en el mapa (Figura 4.18) la dirección de máximo riesgo está posiblemente condicionada por los montes y el cauce del río.

Por último, una aplicación a la mortalidad por cáncer de mama en mu-

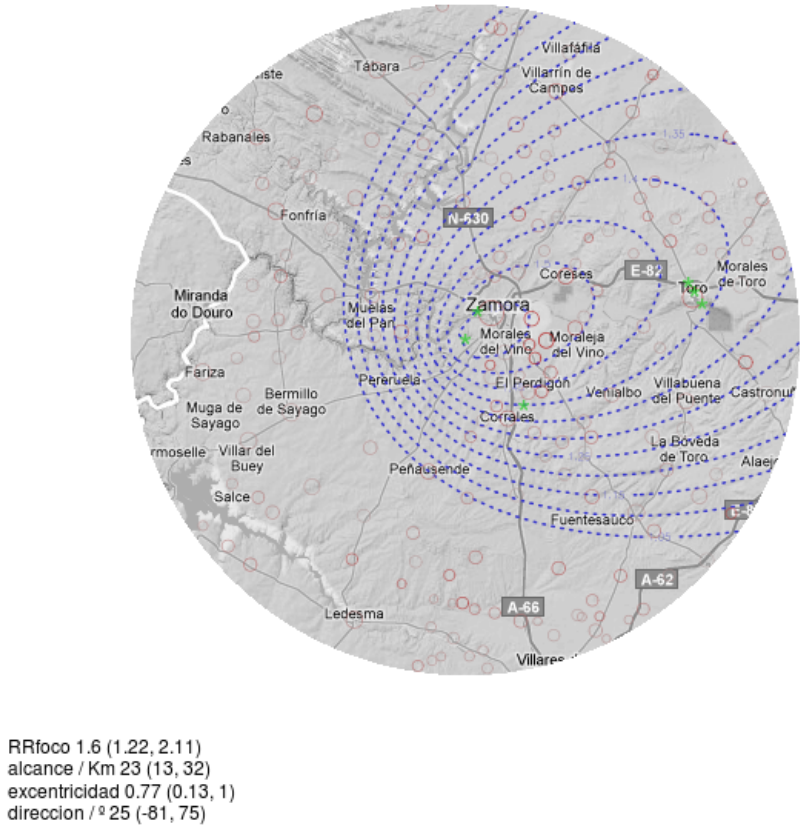


Figura 4.16: Riesgo de morir por cáncer de próstata en el entorno de una industria de Zamora (hombres). Mapa topográfico

Parámetro	**Estimación**	**Lim. Inf. 95 %**	**Lim. Sup. 95 %**
RR_{foco}	2.29	1.69	30.9
Alcance /Km	17	12	21
Excentricidad	0.93	0.78	1.00
Dirección / ^{o}	19	-4	37

Tabla 4.9: Parámetros del modelo anisotropo con umbral para un entorno de Belorado (mujeres). Estimaciones puntuales e intervalo de confianza al 95 %

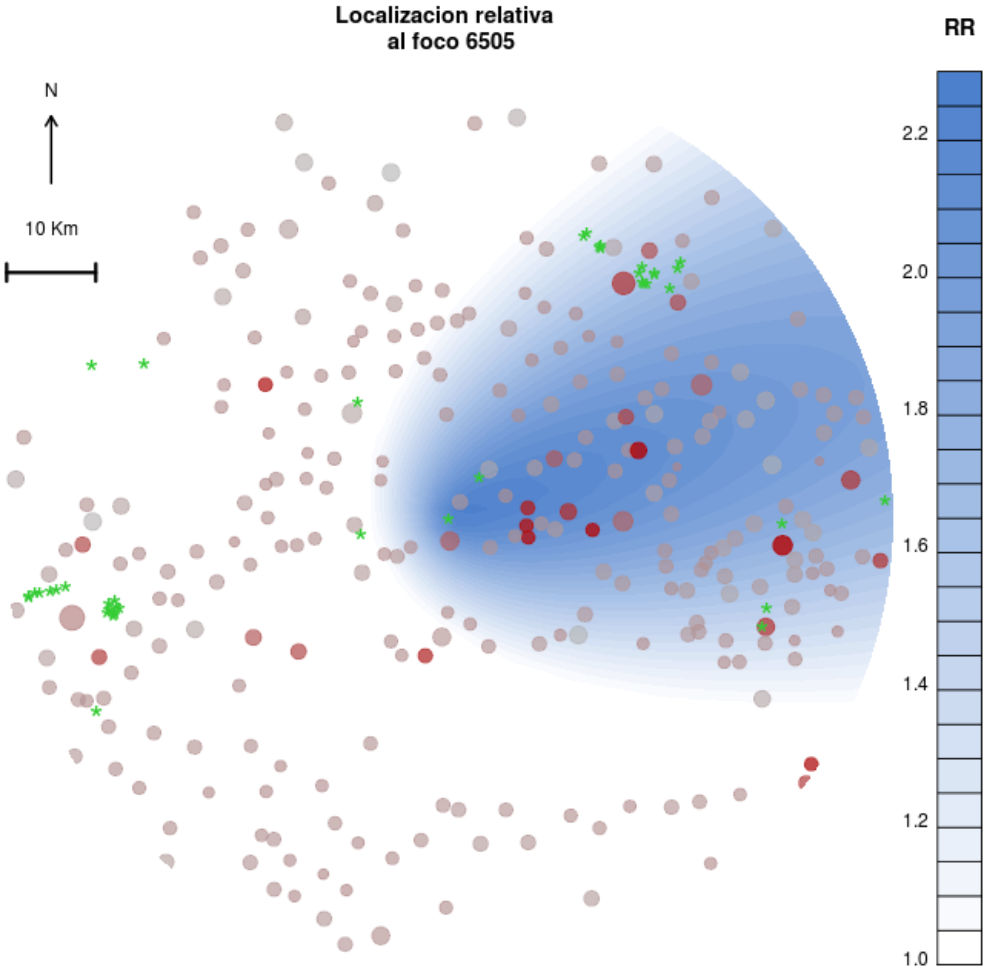


Figura 4.17: Riesgo de morir por cáncer de páncreas en el entorno de una industria de Belorado (mujeres). Gradiente de riesgo

jeres. Se trata del entorno de una industria química en Guareña (Badajoz). La dirección de máximo riesgo estimada apunta hacia el embalse de Alange (Figura 4.20). La Tabla 4.10 contiene las estimaciones de los parámetros del modelo ajustado.

Como información complementaria a estos análisis, la magnitud de las emisiones de los contaminantes más relevantes (arsénico, cadmio, cloro, co-

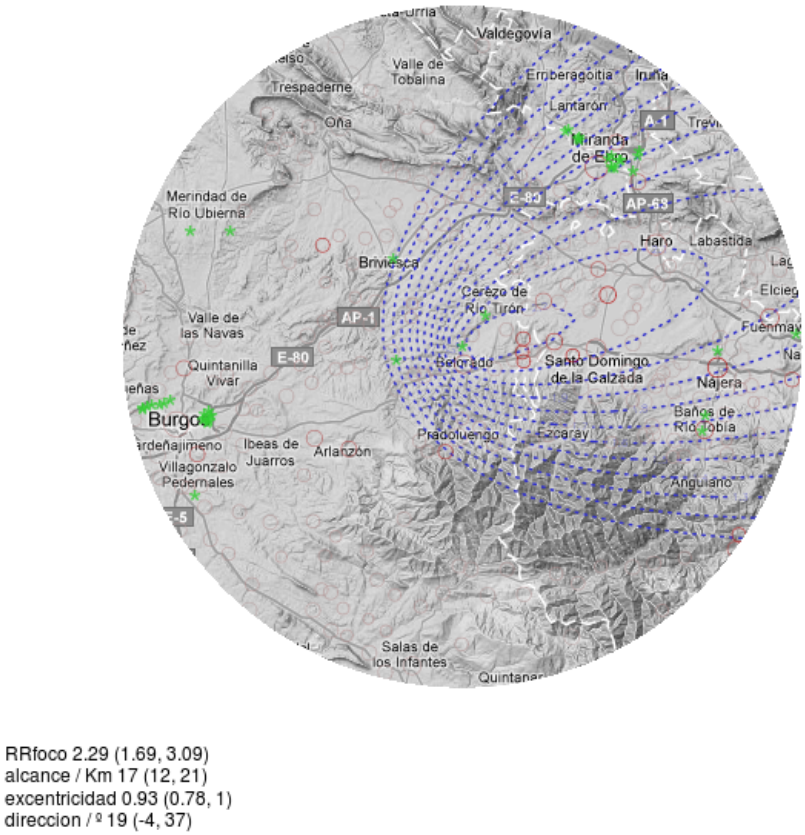


Figura 4.18: Riesgo de morir por cáncer de páncreas en el entorno de una industria de Belorado (mujeres). Mapa topográfico

Parámetro	**Estimación**	**Lim. Inf. 95 %**	**Lim. Sup. 95 %**
RR_{foco}	1.68	1.03	2.74
Alcance /Km	23	8	34
Excentricidad	0.8	0	1.00
Dirección / ^{o}	239	176	313

Tabla 4.10: Parámetros del modelo anisotropo con umbral para un entorno de Guareña (mujeres). Estimaciones puntuales e intervalo de confianza al 95 %

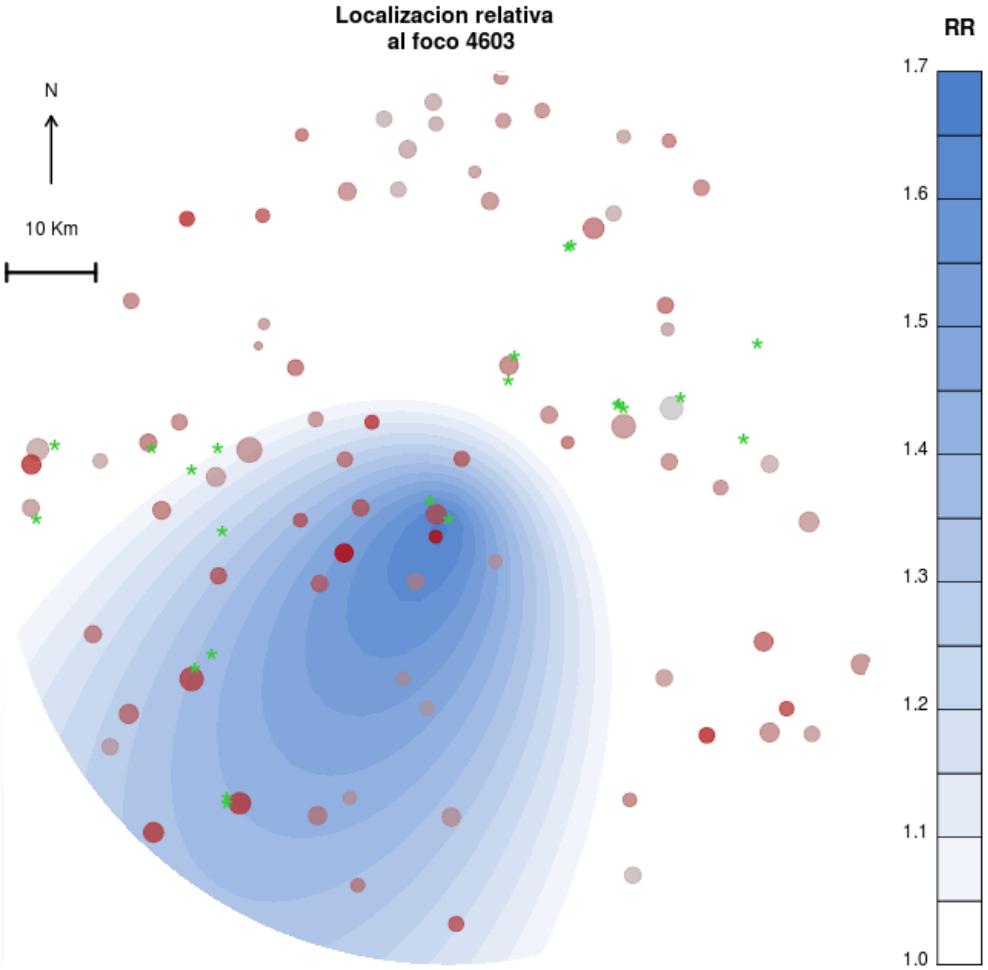
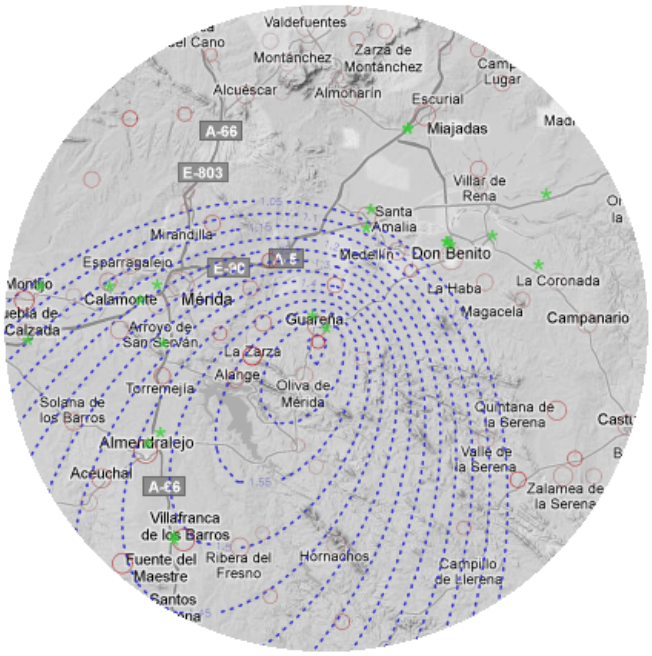


Figura 4.19: Riesgo de morir por cáncer de mama en el entorno de una industria de Guareña (mujeres). Gradiente de riesgo

bre, cromo, dióxido de carbono, flúor, hidroclorofluorocarburos, hidrofluorocarburos, mercurio, monóxido de carbono, níquel, óxido nitroso, óxidos de azufre, óxidos de nitrógeno, partículas, dioxinas y furanos, plomo, tricloroetileno, vanadio, zinc) por cada una de las industrias estudiadas se presenta (en kilogramos anuales) en la Tabla 4.11.



RRfoco 1.68 (1.03, 2.74)
alcance / Km 23 (8, 34)
excentricidad 0.8 (0, 1)
direccion / º 239 (176, 313)

Figura 4.20: Riesgo de morir por cáncer de mama en el entorno de una industria de Guareña (mujeres). Mapa topográfico

Municipio	Tipo	As	Cd	Cl	Cu	Cr	F-Carburos	F	Cl-Carburos	CO2	Hg
Girona	alimentación	0.15	0.83	0	0	1.05	103 10^6	0	0	0	0
Miranda	papel	0	0	0	0	0	46 10^6	0	0	0	0
As Pontes	combustión	525	20.4	26 10^4	657	208	8920 10^6	32 10^4	600	12.9	353
Zamora	residuos	0	0	0	0	0	94 10^4	0	0	0	0
A Pontenova	metal	0	0	0	0	0	2,7 10^4	0	0	0	0
Guareña	química	0.04	0.03	0	0.64	0.03	181 10^4	0	0	0	0.03
Belorado	mineral	0	0	0	0	0	82 10^6	0	0	0	0
Ronda	química	0	0	0	0	0	0	0	0	0	0
Manilva	mineral	0	0	0	0	0	0	0	0	0	0
Girona	residuos	0	0	0	0	0	476 10^4	0	0	0	0

Municipio	CO	Ni	N2O	SOx	NOx	PM10	PCDD	Pb	Tricloroetileno	V	Zn
Girona	6,3 10^4	0	1660	0	7,5 10^4	5720	0	0	0	0	0
Miranda	96 10^4	0	0	16 10^4	27 10^4	1,6 10^4	0	0	0	0	0
As Pontes	45 10^4	186	4,7 10^4	211 10^6	14 10^4	1,7 10^6	0	183	0	0	1870
Zamora	0	0	0	3,2 10^6	7 10^4	0	0	0	0	0	0
A Pontenova	7.21	0	0	7.82	27.2	0	0	0	0	0	0
Guareña	296	0.03	387	2330	2220	0	0	0.10	0	0.64	0.04
Belorado	1,1 10^4	0	0	17 10^4	26 10^4	0	0	0	0	0	0
Ronda	0	0	0	0	0	0.01	0	0	20	0	0
Manilva	0	0	0	0	0	3200	0	0	0	0	0
Girona	0	0	0	0	3510	0	0	0	0	0	0

Tabla 4.11: Emisiones declaradas en el año 2007 por las industrias estudiadas (en Kg al año)

CAPÍTULO 5

Discusión

"Las nubes no son esferas, las montañas no son conos, las costas no son circulares y las cortezas de los árboles no son lisas, así como los relámpagos no viajan en línea recta"

Benoît Mandelbrot

5.1. Aportaciones metodológicas

Existen escasas propuestas que tengan en cuenta la componente direccional a la hora de estudiar la distribución de las enfermedades en los entornos de focos contaminantes (Lawson 1993; Congdon 2003). En ellas, el tratamiento aplicado a las variables direccionales es similar al de una variable confusora, en cuanto a que buscan controlar su efecto, pero no estimarlo ni contrastar la existencia de manera directa. En esas propuestas, la introducción de términos angulares en la regresión se realiza en base a su significati-

vidad, restando coherencia al modelo y sin tener en cuenta la interpretación del resultado final. Las aplicaciones presentadas en esas referencias hacen uso de una parametrización de los modelos que no incluye los términos necesarios para una completa descripción espacial direccional, con lo que no resulta viable una interpretación directa ni se cuantifica correctamente la anisotropía modelizada. Además, no todas contemplan la combinación con un factor umbral en el alcance.

De manera original, esta tesis aporta nuevas técnicas espaciales para estudiar los entornos de focos contaminantes. En concreto, los modelos propuestos son capaces de detectar un umbral de alcance y/o una dirección en la asociación espacial, haciendo uso de modelos generalizados con parametrizaciones anidadas. Estos modelos se han desarrollado de manera coherente, teniendo siempre presente la interpretación y comparación de los resultados.

Las nuevas metodologías presentadas y desarrolladas a lo largo de este trabajo superan algunas de las carencias de las, apuntadas en la Sección 1.4 del Capítulo 1:

- No dependen de decisiones arbitrarias a priori, como los puntos de corte para los modelos categóricos, el orden del polinomio en los modelos polinomiales o el grado de suavización en los modelos no paramétricos
- Al ser modelos de regresión, superan a los test no paramétricos de asociación en capacidad de ajuste por covariables así como cuantificación de las estimaciones del riesgo
- La información de la que dependen es limitada (localización del foco y de las áreas de estudio) en comparación con los modelos que se basan

en estimaciones del nivel de contaminantes en cada población (como los modelos de dispersión atmosférica, que requieren gran cantidad de información del entorno estudiado)

- Por contraposición a la metodología no paramétrica, los resultados de la estimación cuantifican de manera directa la magnitud del riesgo y su variabilidad
- La forma de la distribución del riesgo modelizada no exhibe saltos abruptos (como modelos categóricos) y ni grandes problemas de plausibilidad epidemiológica (como los riesgos infinitos en el foco o a grandes distancias de los modelos lineales o logarítmicos)

En cuanto a la implementación de las propuestas, se ha dado con procedimientos fiables, rápidos y eficaces para superar las limitaciones que la estimación de la variabilidad y la comparación de modelos presentan. Al realizar la primera basándose en la verosimilitud, se evitan los procedimientos de Monte Carlo que tienen un coste computacional mayor. Este ahorro tiene su importancia en la aplicación sistemática para la exploración de grandes conjuntos de datos.

La comparación entre los modelos, posible al tratarse de parametrizaciones anidadas, es de suma importancia, ya que permite establecer pruebas de existencia de las distintas componentes espaciales, creando un "protocolo de decisión" para caracterizar el riesgo en los escenarios a estudiar y facilitando la interpretación de los modelos.

De manera general, los modelos propuestos hacen un uso más eficiente de la localización geográfica en estudios de entornos de focos contaminantes,

aumentando la capacidad de detección de asociaciones espaciales con el foco.

5.2. Discusión sobre las aplicaciones

La eficiencia computacional de los procedimientos permite su utilización para una exploración sistemática de grandes conjuntos de datos. Como se puede comprobar en la Sección 4.2 del Capítulo 4, las nuevas propuestas aumentan considerablemente la capacidad de detectar efectos espaciales relacionados con los presuntos focos contaminantes.

Se han estudiado con más detalle localizaciones en las que existe un riesgo de morir por distintos tipos de cáncer asociado a alguno de los focos industriales estudiados. Las características espaciales de esas localizaciones invitan a pensar que orografía y meteorología están implicadas en la distribución del riesgo.

El escenario estudiado en Girona muestra una distribución espacial de morir por cáncer de estómago similar en ambos sexos, con una direccionalidad acusada hacia el oeste. Las estimaciones de los parámetros son compatibles entre los dos sexos. Este hecho indica la existencia de una posible causa ambiental. Por otra parte, la proximidad de los dos focos industriales existentes en la ciudad no permite discernir cual de ellos pueda ser el causante.

La mortalidad por cáncer de vejiga en hombres en los alrededores de la industria pirotécnica en Ronda sigue patrón direccional que apunta en la dirección contraria a la sierra que se encuentra al sur. Este accidente

geográfico podría estar relacionado con la distribución atmosférica de los contaminantes emitidos por el foco.

El cauce del río Ebro parece marcar la dirección de máximo riesgo de morir por cáncer de vejiga con respecto a una industria papelera de Miranda de Ebro, aunque la presencia de una importante vía de comunicación, la autopista A-68 hacia Logroño, podría condicionar el patrón de mortalidad detectado.

Existe un exceso de riesgo de morir por cáncer colorrectal en las mujeres residentes entre las provincias de A Coruña y Lugo, centrada en la comarca Terra Chá. El modelo anisótropo con umbral detecta una asociación de esta causa con dos focos contaminantes: La central Térmica de As Pontes y una industria metalúrgica en A Pontenova. La zona con incremento de riesgo es una planicie rodeada por un sistema montañoso (Montes Litorales), que podría estar condicionando la meteorología y el patrón de dispersión de los contaminantes.

La mortalidad por cáncer de pulmón en mujeres en la provincia de Málaga parece estar asociada a la cantera de Manilva, aunque el patrón detectado es compatible con un aumento de riesgo debido a la proximidad del mar.

En cuanto a la distribución de la mortalidad por cáncer de próstata en Zamora, la dirección de máximo riesgo detectada coincide con el cauce de río y con la dirección del viento predominante en la zona.

De nuevo la cuenca fluvial y los montes pueden estar implicados en la distribución espacial de la mortalidad por cáncer de páncreas en mujeres

asociado a una cantera en Belorado.

La detección de una direccionalidad en el riesgo de morir por cáncer de mama en mujeres en el entorno de una industria química en Badajoz apunta hacia el embalse de Alange. Sin embargo, a la hora de utilizar este resultado con fines etiológicos, hay que tener en cuenta la alta supervivencia en este tipo de tumores.

5.3. Limitaciones

En primer lugar, no hay que olvidar que se trata de nuevas propuestas para regresiones ecológicas, que heredan todas las limitaciones de este tipo de estudios (Greenland and Morgenstern 1989; Diggle and Elliott 1995; Elliott et al. 2001); trasladar los resultados del nivel de agregación estudiado (municipal en las aplicaciones mostradas) al nivel individual no siempre será posible.

La aparente complejidad de las propuestas no lo es tanto. Basta recordar que se está modelizando el riesgo mediante figuras geométricas simples (círculos y elipses). Esta simplicidad ayuda a la hora de estimar e interpretar, pero impone unas restricciones que limitan la validez de resultados y conclusiones. Los modelos propuestos pretenden ser una estimación global de la distribución espacial del riesgo y no permiten descripciones tan versátiles como los modelos no paramétricos.

Por otra parte, en cada entorno estudiado sólo se tiene en cuenta un posible foco contaminante. La existencia de múltiples focos contaminantes puede

condicionar la distribución del riesgo de manera que reste plausibilidad a los resultados de los modelos propuestos (Ramis et al. 2011). Además, en las situaciones en que dos o más focos estén muy próximos, sus asociaciones espaciales al nivel de agregación estudiado son indistinguibles. Al modelizarse el número de casos mediante una distribución de Poisson podría darse la situación en la que existiera sobredispersión en los datos, es decir, que la variabilidad real de los datos fuera mayor que la asumida por el modelo. Esto resultaría en una infraestimación de la variabilidad de los estimadores, reduciendo la cobertura real de los intervalos de confianza. Se volverá sobre este punto en la sección siguiente.

Si bien la mortalidad es el mejor indicador disponible con la cobertura y el nivel de agregación requeridos, está condicionado por la supervivencia a la hora de sacar conclusiones etiológicas. Esto puede introducir un sesgo de selección a la hora de buscar explicaciones etiológicas a los resultados en el caso de que la distribución espacial de la supervivencia (para cada causa) no sea homogénea. Es decir, hay que tener presente que los resultados de las aplicaciones del Capítulo 4 son válidos para la mortalidad y es necesario tener precaución a la hora de extrapolarlos en busca de posibles causantes de la enfermedad.

En cuanto al sesgo de información, el registro de los focos contaminantes puede no ser del todo fiable. Las localizaciones geográficas (García-Pérez et al. 2008) y emisiones de las industrias del registro E-PRTR son, en algunos casos, imprecisas o incorrectas. Un error de unos pocos kilómetros en la situación del foco contaminante cambia drásticamente el patrón estimado

por los modelos, pudiendo invalidar la posible asociación con el foco y, por tanto, los resultados de los análisis.

También pueden existir problemas a la hora de comparar la mortalidad entre las distintos municipios debido a desigualdades no tenidas en cuenta (sesgo de confusión). Por una parte, pueden existir factores confusores no controlados en el análisis (tabaco, por ejemplo). Además, es posible que la relación de la mortalidad con las variables socioeconómicas sea más compleja que la asumida en los modelos. El segundo problema se puede paliar mediante modelizaciones más flexibles de estas variables y el primero, obteniendo más información.

La presencia de grandes núcleos poblacionales puede condicionar la direccionalidad de los escenarios. Sin embargo, el análisis de sensibilidad llevado a cabo eliminándolos permite aumentar confianza en los resultados obtenidos.

5.4. Posibles ampliaciones

Podría mejorarse la validez de los resultados incluyendo explícitamente en el modelo más de un foco. Esto permitiría el estudio de zonas densamente industrializadas, que suelen ser también las más pobladas. Aun así, incluir una variable radial y otra angular para cada uno de los múltiples focos es complejo a todos los niveles: desde la modelización hasta la comparación de modelos, pasando por la estimación de la variabilidad de los parámetros, sin mencionar la dificultad para aislar en interpretar las contribuciones de los distintos focos en un escenario con varios de ellos.

Tener en cuenta estructuras espaciales más complejas que las propuestas puede realizarse mediante enfoques no paramétricos. Habría que desarrollar métodos específicos para realizar inferencia sobre esta metodología principalmente descriptiva.

Una ampliación más sencilla que tal vez mejoraría los resultados consiste en incluir información a cerca de la altitud de las áreas estudiadas con respecto al foco. Con esto se tendría control sobre posibles efectos en la dispersión atmosférica y, sobre todo, en la dispersión fluvial (río abajo). El tratamiento de la nueva variable altimétrica podría hacerse mediante alguna de las parametrizaciones expuestas en los antecedentes metodológicos o bien combinarse con la información existente (distancia y ángulo).

Implementar en los modelos distribuciones de parámetro de dispersión libre (quasi-Poisson, binomial negativa) permitiría estimar de manera más adecuada la variabilidad en las situaciones en las que exista sobredispersión en los datos.

Si esta metodología se aplica a indicadores de salud de respuesta rápida (rápida en contraposición con el cáncer, por ejemplo, que presenta largos periodos de latencia), podría incluirse también información temporal a cerca de los periodos de emisión por parte del foco contaminante.

Con respecto a la aplicación práctica, estudiar la incidencia en vez de la mortalidad mejoraría la interpretación etiológica de los resultados. Es de esperar que actualizaciones sucesivas del registro de contaminantes (E-PRTR) mejoren la calidad de la información a cerca de los focos contaminantes. En todo caso, estudios detallados de las zonas sospechosas podrían arrojar luz

a cerca del efecto concreto de cada foco.

CAPÍTULO 6

Resumen

A lo largo de este trabajo se han motivado, propuesto, desarrollado, implementado, comprobado y aplicado nuevas metodologías para el estudio de la posible relación entre un foco contaminante y la enfermedad en la población residente en su entorno.

Desde el punto de vista de la estadística espacial se han estudiado dos aspectos: La existencia de un alcance y de una direccionalidad en la asociación espacial con el foco. El abordaje de estas cuestiones se ha hecho tanto de manera independiente como conjunta, llegando al enunciado de tres modelos espaciales: El modelo radial con umbral (caracterizado por $f_{radial+umbral}$, ecuación 3.4), el modelo anisótropo ($f_{anisotropo}$, ecuación 3.6) y el modelo anisótropo con umbral ($f_{anisotropo+umbral}$, ecuación 3.7). Los detalles de la motivación y parametrización de estos modelos se pueden seguir el Capítulo 3 y, con más profundidad, en Apéndice A.

Los procedimientos para estimar los parámetros y sus variabilidades co-

rrespondientes, en cada uno de los modelos, así como la comparación entre ellos, se detallan en el Apéndice A y han sido puestos a prueba en el estudio de simulación (Apéndice B).

Una vez elegidas las técnicas óptimas para ajustar y comparar los modelos, se han aplicado a escenarios con datos de mortalidad por las causas tumorales más frecuentes en el entorno de las industrias del registro E-PRTR, como se muestra en el Capítulo 4.

CAPÍTULO 7

Conclusiones

1. En los entornos de focos contaminantes se pueden estimar características de una supuesta relación espacial del riesgo con el foco emisor. En concreto: detectar un umbral en el alcance de la asociación, modelizar una distribución angular alrededor de una dirección de máximo riesgo y combinar estos dos factores.

2. La estructura anidada de las parametrizaciones espaciales propuestas (modelo nulo, radial, radial con umbral, anisótropo y anisótropo con umbral) permite el planteamiento de pruebas de existencia de las distintas componentes espaciales del riesgo.

3. El rendimiento de los métodos asintóticos basados en la verosimilitud presentados es satisfactoria. Se pueden utilizar para estimar la variabilidad de los parámetros y realizar la comparación de modelos.

4. Los dos puntos anteriores permiten establecer unas reglas de decisión encadenadas que caractericen la asociación espacial presente en cada

localización.

5. La eficacia y rapidez de la metodología propuesta permite su uso de manera sistemática para realizar análisis exploratorios en busca de localizaciones susceptibles de presentar una dependencia de los indicadores de enfermedad con la presencia de focos contaminantes.

6. Las simplificaciones en la modelización y la presencia de múltiples focos contaminantes limitan en parte la generalización de los resultados.

7. Con respecto a la aplicación práctica estudiada, algunas de las industrias presentes en el registro E-PRTR presentan relación espacial con la mortalidad por diversas causas tumorales en sus entornos. Estos resultados están condicionados por las limitaciones expuestas y las propias de un estudio ecológico.

Bibliografía

H-O. Adami, D. Hunter, and D. Trichopoulos. *Textbook of Cancer Epidemiology*. Oxford University Press, USA, 2008.

M. Benedetti, I. Iavarone, P. Comba, and I. Lavarone. Cancer risk associated with residential proximity to industrial sites: a review. *Archives of Environmental Health*, 56, 2001.

Y. Benjamini. The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29, 2001.

J. Besag and J. Newell. The detection of clusters in rare diseases. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 154, 1991.

A. Biggeri, F. Barbone, C. Lagazio, M. Bovenzi, and G. Stanta. Air pollution and lung cancer in trieste, italy: spatial analysis of risk as a function of distance from sources. *Environmental Health Perspectives*, 104, July 1996.

J. F. Bithell. The choice of test for detecting raised disease risk near a point source. *Statistics in Medicine*, 14, 1995.

R. S. Bivand, E. J. Pebesma, and V. Gómez-Rubio. *Applied Spatial Data Analysis with R*. Springer, 1 edition, 2008.

A. W. Bowman and A. Azzalini. *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations*. Oxford University Press, 1997.

N. E. Breslow, J. H. Lubin, P. Marek, and B. Langholz. Multiplicative models and cohort analysis. *Journal of the American Statistical Association*, 78, 1983.

N. E. Breslow, N. E. Day, and International Agency for Research on Cancer. *Statistical Methods in Cancer Research: The Analysis of Case-control Studies Vol 1*. International Agency for Research on Cancer, 1993.

D. Clayton and M. Hills. *Statistical Models in Epidemiology*. Oxford University Press, USA, 1993.

P. Congdon. *Applied bayesian modelling*. John Wiley & Sons Inc, 2003.

J. Cuzick and R. Edwards. Spatial clustering for inhomogeneous populations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 52, 1990.

P. Diggle and P. Elliott. Disease risk near point sources: statistical issues for analyses using individual or spatially aggregated data. *Journal of Epidemiology and Community Health*, 49 Suppl 2, 1995.

P. Diggle, S. Morris, P. Elliott, and G. Shaddick. Regression modelling of disease risk in relation to point sources. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 160, 1997.

R. Edwards, T. Pless-Mulloli, D. Howel, t. Chadwick, R. Bhopal, R. Harrison, and H. Gribbin. Does living near heavy industry cause lung cancer in women? a case-control study using life grid interviews. *Thorax*, 61, 2006.

B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap*. Chapman and Hall/CRC, 1994.

P. Elliott and D. Wartenberg. Spatial epidemiology: current approaches and future challenges. *Environmental health perspectives*, 112, 2004.

P. Elliott, M. Hills, J. Beresford, I. Kleinschmidt, D. Jolley, S. Pattenden, L. Rodrigues, A. Westlake, and G. Rose. Incidence of cancers of the larynx and lung near incinerators of waste solvents and oils in great britain. *Lancet*, 339, 1992.

P. Elliott, J. Cuzick, D. English, and R. Stern. *Geographical and Environmental Epidemiology: Methods for Small-Area Studies*. Oxford University Press, USA, 1996.

P. Elliott, J. Wakefield, N. Best, and D. Briggs. *Spatial Epidemiology: Methods and Applications*. Oxford University Press, 2001.

J. Ferlay, H-R. Shin, F. Bray, D. Forman, C. Mathers, and D. M. Parkin. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *International Journal of Cancer. Journal International Du Cancer*, 2010.

J. García-Pérez, E. Boldo, R. Ramis, E. Vidal, N. Aragonés, B. Pérez-Gómez, M. Pollán, and G. López-Abente. Validation of the geographic position of EPER-Spain industries. *International Journal of Health Geographics*, 7, 2008.

J. García-Pérez, M. Pollán, E. Boldo, B. Pérez-Gómez, N. Aragonés, V. Lope, R. Ramis, E. Vidal, and G. López-Abente. Mortality due to lung, laryngeal and bladder cancer in towns lying in the vicinity of combustion installations. *The Science of the Total Environment*, 407, 2009.

J. García-Pérez, M. F. López-Cima, B. Pérez-Gómez, N. Aragonés, M. Pollán, E. Vidal, and G. López-Abente. Mortality due to tumours of the digestive system in towns lying in the vicinity of metal production and processing installations. *The Science of the Total Environment*, 408, 2010.

C. A. Gonzalez, M. Kogevinas, E. Gadea, A. Huici, A. Bosch, M. J. Bleda, and O Päpke. Biomonitoring study of people living near or working at a municipal solid-waste incinerator before and after two years of operation. *Archives of Environmental Health*, 55, 2000.

S. Goria, C. Daniau, P. de Crouy-Chanel, P. Empereur-Bissonnet, P. Fabre, M. Colonna, C. Duboudin, J-F. Viel, and S. Richardson. Risk of cancer in the vicinity of municipal solid waste incinerators: importance of using a flexible modelling strategy. *International Journal of Health Geographics*, 8, 2009.

M. S. Gottlieb, C. L. Shear, and C. L. Seale. Lung cancer mortality and residential proximity to industry. *Environmental Health Perspectives*, 45, 1982.

S. Greenland. Dose-response and trend analysis in epidemiology: alternatives to categorical analysis. *Epidemiology*, 6, 1995.

S. Greenland and H. Morgenstern. Ecological bias, confounding, and effect modification. *International Journal of Epidemiology*, 18, 1989.

J. O. Grimalt, J. Sunyer, V. Moreno, O. C. Amaral, M. Sala, A. Rosell, J. M. Anto, and J. Albaiges. Risk excess of soft-tissue sarcoma and thyroid cancer in a community exposed to airborne organochlorinated compound

mixtures with a high hexachlorobenzene content. *International Journal of Cancer. Journal International Du Cancer*, 56, 1994.

T. J. Hastie and R. J. Tibshirani. *Generalized Additive Models*. Chapman and Hall/CRC, 1990.

K. C. Johnson, S. Pan, R. Fry, and Y Mao. Residential proximity to industrial plants and non-Hodgkin lymphoma. *Epidemiology (Cambridge, Mass.)*, 14, 2003.

A. B. Lawson. On the analysis of mortality events associated with a prespecified fixed point. *Journal of the Royal Statistical Society. Series A, (Statistics in Society)*, 156, 1993.

A. B. Lawson, A. Biggeri, D. Böhning, E. Lesaffre, J-F. Viel, and R. Bertollini. *Disease Mapping and Risk Assessment for Public Health*. Wiley, 1999.

A. B. Lawson, W. J. Browne, and C. L. Vidal Rodeiro. *Disease Mapping with WinBUGS and MLwiN*. Wiley, 2003.

A. Linos, A. Blair, R. W. Gibson, G. Everett, S. Van Lier, K. P. Cantor, L. Schuman, and L. Burmeister. Leukemia and non-Hodgkin's lymphoma and residential proximity to industrial plants. *Archives of Environmental Health*, 46, 1991.

G. López-Abente, N. Aragonés, M. Pollán, M. Ruiz, and A. Gandarillas. Leukemia, lymphomas, and myeloma mortality in the vicinity of nuclear power plants and nuclear fuel facilities in spain. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association*

for Cancer Research, Cosponsored by the American Society of Preventive Oncology, 8, 1999.

G. López-Abente, N. Aragonés, and M. Pollán. Solid-tumor mortality in the vicinity of uranium cycle facilities and nuclear power plants in spain. *Environmental Health Perspectives*, 109, 2001.

G. López-Abente, R. Ramis, M. Pollán, N. Aragonés, B. Pérez-Gómez, D. Gómez-Barroso, J-M. Carrasco, V. Lope, J. García-Pérez, E. Boldo, and M-J. García-Mendizábal. *Atlas municipal de mortalidad por cáncer en España 1989-1998*. Instituto de Salud Carlos III, 2007.

C. Magnani, C. Agudo, C. A. González, A. Andrion, A. Calleja, E. Chellini, P. Dalmasso, A. Escolar, S. Hernandez, C. Ivaldi, D. Mirabelli, J. Ramirez, D. Turuguet, M. Usel, and B. Terracini. Multicentric study on malignant pleural mesothelioma and non-occupational exposure to asbestos. *British Journal of Cancer*, 83, 2000.

P. McCullagh and J. A. Nelder. *Generalized Linear Models, Second Edition*. Chapman and Hall/CRC, 1989.

P. Michelozzi, D. Fusco, F. Forastiere, C. Ancona, c. Dell'Orco, and C. A. Perucci. Small area study of mortality among people living near multiple sources of air pollution. *Occupational and Environmental Medicine*, 55, 1998.

T. Morton-Jones, P. Diggle, and P. Elliott. Investigation of excess environmental risk around putative sources: Stone's test with covariate adjustment. *Statistics in Medicine*, 18, 1999.

J. N. Newell and J. E. Besag. Methods for investigating localized clustering of disease. the detection of small-area database anomalies. *IARC Scientific Publications*, 1996.

S. Openshaw, A. W. Craft, M. Charlton, and J. M. Birch. Investigation of leukaemia clusters by use of a geographical analysis machine. *Lancet*, 1, 1988.

D. Ozalla, C. Herrero, N. Ribas-Fitó, J. To-Figueras, A. Toll, M. Sala, J. Grimalt, X. Basagaña, M. Lecha, and J. Sunyer. Evaluation of urinary porphyrin excretion in neonates born to mothers exposed to airborne hexachlorobenzene. *Environmental Health Perspectives*, 110, 2002.

S. Parodi, E. Stagnaro, C. Casella, A. Puppo, E. Daminelli, V Fontana, F Valerio, and M. Vercelli. Lung cancer in an urban area in northern italy near a coke oven plant. *Lung Cancer (Amsterdam, Netherlands)*, 47, 2005.

J. Pekkanen, E. Pukkala, M. Vahteristo, and T. Vartiainen. Cancer incidence around an oil refinery as an example of a small area study based on map coordinates. *Environmental Research*, 71, 1995.

R. Ramis, P. Diggle, H. Cambra, and G. López-Abente. Prostate cancer and industrial pollution risk around putative focus in a multi-source scenario. *Environment International*, 2011.

K.J. Rothman, S. Greenland, and T.L. Lash. *Modern epidemiology*. Lippincott Williams & Wilkins, 2008.

S. Sans, P. Elliott, I. Kleinschmidt, G. Shaddick, S. Pattenden, P. Walls, C. Grundy, and H. Dolk. Cancer incidence and mortality near the baglan

bay petrochemical works, south wales. *Occupational and Environmental Medicine*, 52, 1995.

G. A. F. Seber and C. J. Wild. *Nonlinear Regression*. John Wiley & Sons Canada, Ltd., 1989 edition, 1989.

G. Shaddick and P. Elliott. Use of stone's method in studies of disease risk around point sources of environmental pollution. *Statistics in Medicine*, 15, 1996.

A. Silva-Mato, D. Viana, M. I. Fernández-SanMartín, J. Cobos, and M. Viana. Cancer risk around the nuclear power plants of trillo and zorita (Spain). *Occupational and Environmental Medicine*, 60, 2003.

R. A. Stone. Investigations of excess environmental risks around putative sources: statistical problems and a proposed test. *Statistics in Medicine*, 7, 1988.

J. Sunyer, C. Herrero, D. Ozalla, M. Sala, N. Ribas-Fitó, J. Grimalt, and X. Basagaña. Serum organochlorines and urinary porphyrin pattern in a population highly exposed to hexachlorobenzene. *Environmental Health: A Global Access Science Source*, 1, 2002.

M. Szklo and F.J. Nieto. *Epidemiology: beyond the basics*. Jones & Bartlett Learning, 2007.

R. E. Tarone. The use of historical control information in testing for a trend in poisson means. *Biometrics*, 38, 1982.

S. Urbanek. *multicore: Parallel processing of R code on machines with multiple cores or CPUs*, 2009.

W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S.* Springer, 2002.

J-F. Viel, C. Daniau, D. Goria, P. Fabre, P. de Crouy-Chanel, E-A. Sauleau, and P. Empereur-Bissonnet. Risk for non hodgkin's lymphoma in the vicinity of french municipal solid waste incinerators. *Environmental Health: A Global Access Science Source*, 7, 2008.

L. A. Waller and C. A. Gotway. *Applied Spatial Statistics for Public Health Data.* Wiley-Interscience, 2004.

L. A. Waller and A. B. Lawson. The power of focused tests to detect disease clustering. *Statistics in Medicine*, 14, 1995.

H. Wickham. *ggplot2: Elegant Graphics for Data Analysis.* Springer, 2009.

Área de Epidemiología Ambiental y cáncer Centro Nacional de Epidemiología ISCIII. Mortalidad por cáncer y otras causas en España, 2008.

APÉNDICE A

Parametrización e implementación

A.1. Parametrización

La notación y consideraciones generales sobre el modelo son las mismas que las expuestas en la Sección 3.1 de la página 36.

Modelo radial

El punto de partida de las nuevas parametrizaciones es el modelo radial, en el que se asume una dependencia lineal de la función espacial con la distancia (Ecuación A.1).

$$f_{radial}(d_i|\beta_0,\beta_1) = \beta_0 + \beta_1\left(1 - \frac{d_i}{d_{ref}}\right) = \beta_0' + \beta_1' d_i \tag{A.1}$$

En donde d_{ref} es la distancia de referencia, de la que se puede obtener la RME asociada exponenciando el coeficiente β_0. El otro coeficiente (β_1) se corresponde con logaritmo del riesgo relativo en el foco con respecto a la distancia de referencia.

La función definida en la Ecuación A.1 es lineal en los parámetros β'_0 y β'_1, con lo que su estimación es directa. A partir de ellos, se pueden recuperar los parámetros de interés mediante las transformaciones siguientes

$$\beta_0 = \beta'_0 + d_{ref}\beta'_1$$
$$\beta_1 = -d_{ref}\beta'_1$$

Modelo radial con umbral

Como extensión al modelo radial y para superar la limitación que presenta a grandes distancias, se plantea la posibilidad de modelizar la distancia de tal manera que exista una relación proporcional con el riesgo (lineal) hasta un determinado umbral a partir del cual no exista asociación. Esto es, limitar el modelo radial a un rango acotado, estimando tanto la tasa de proporcionalidad del riesgo como la distancia a la que deja de existir asociación.

Para ello se propone la siguiente parametrización, en donde $(x)^+$ se define

igual que en la ecuación 3.5:

$$f_{radial+umbral}(d_i|\beta_0, \beta_1, \lambda) = \beta_0 + \beta_1 \left(1 - \frac{d_i}{\lambda}\right)^+ \tag{A.2}$$

En este caso, λ es la distancia que define el umbral de alcance de la asociación, β_0 se corresponde con logaritmo de la RME en la zona de referencia y β_1 con el logaritmo del riesgo relativo en el foco respecto a dicha zona.

El riesgo evoluciona de manera lineal con la distancia dentro de un circulo de radio λ. Fuera de él se considera un riesgo constante no asociado con la posición.

Modelo anisótropo

Partiendo de uno de los modelos más simples (el radial), se trata de generalizar la parametrización de manera que permita dependencia angular en el riesgo (anisotropía espacial). Es una transición de un modelo circular (en el que las lineas isoriesgo son circunferencias concéntricas en el foco) a uno elíptico (lineas isoriesgo elípticas).

Teniendo en cuenta la interpretación de los parámetros de la Ecuación A.1 se busca un modelo que mantenga las características de β_0 y β_1 (RME de referencia y riesgo relativo en el foco), pero en el que la distancia de referencia varíe con el ángulo, describiendo elipses.

La expresión en coordenadas polares de una elipse de excentricidad ϵ,

semi latus rectum [1] d_{ref}, con un foco en el origen de coordenadas y su eje mayor formando un ángulo ω con la horizontal es la siguiente:

$$u(a_i) = \frac{d_{ref}}{1 - \epsilon \cos(a_i - \omega)} \tag{A.3}$$

Al introducir la dependencia angular definida por la ecuación A.3 en el modelo radial (sustituyendo d_{ref} por $u(a_i)$ en la ecuación A.1) se obtiene un modelo con las características deseadas.

$$\begin{aligned} f_{anisotropo}(d_i, a_i|\beta_0, \beta_1, \epsilon, \omega) &= \beta_0 + \beta_1 \left(1 - \frac{d_i}{u(a_i|\epsilon, \omega)}\right) \\ u(a_i|\epsilon, \omega) &= \frac{d_{ref}}{1 - \epsilon cos(a_i - \omega)} \end{aligned} \tag{A.4}$$

En donde

- ϵ = excentricidad
- ω = ángulo del semieje mayor con la horizontal

A destacar que si $\epsilon = 0 \Rightarrow u(a_i|\epsilon, \omega) = d_{ref} \Rightarrow f_{anisotropo} = f_{radial}$, con lo que la parametrización anisótropa queda reducida a la radial.

Con este modelo se obtienen lineas isoriesgo elípticas como se buscaban, pero la expresión de la Ecuación A.4 no es lineal en los parámetros.

Teniendo en cuenta la descomposición del coseno de la suma de ángulos

[1] El semi-latus rectum es la distancia de la elipse a uno de sus focos en la dirección perpendicular al semi eje mayor.

$$\cos(\alpha - \beta) = \sin(\alpha)\sin(\beta) + \cos(\alpha)\cos(\beta)$$

se puede reparametrizar la expresión de dependencia espacial anisótropa de la siguiente manera

$$\begin{aligned} f_{anisotropo} &= \beta_0 + \beta_1 - \frac{\beta_1}{d_{ref}} d_i + \frac{\beta_1 \epsilon}{d_{ref}} \sin(\omega)\sin(a_i) d_i + \frac{\beta_1 \epsilon}{d_{ref}} \cos(\omega)\cos(a_i) d_i \\ &= \beta'_0 + \beta'_1 d_i + \beta_s \sin(a_i) d_i + \beta_c \cos(a_i) d_i \end{aligned}$$

Esta función sí es lineal en los nuevos parámetros (β'_0, β'_1, β_S, β_C), con lo que su estimación se simplifica a un modelo de regresión lineal generalizado. Además, se pueden recuperar los coeficientes iniciales (de interpretación directa) a partir de los nuevos mediante las relaciones

$$\begin{aligned} \beta_0 &= \beta'_0 + d_{ref}\beta'_1 \\ \beta_1 &= -\lambda\beta'_1 \\ \epsilon &= \sqrt{\frac{{\beta_S}^2 + {\beta_C}^2}{{\beta'_1}^2}} \\ \omega &= \operatorname{arc\,tg} \frac{\beta_S}{\beta_C} \end{aligned}$$

Modelo anisótropo con umbral

La función $f_{anisotropo}$ (ecuación A.4) varía con la distancia y el ángulo, pero, al igual que el modelo radial, su comportamiento a grandes distancias no

es satisfactorio. Se puede, igual que se ha hecho con la función $f_{radial+umbral}$ (ecuación A.2), restringir la asociación espacial dentro de un alcance λ. Se trata estimar el valor de d_{ref} mediante el parámetro λ, sin fijar su valor de antemano.

$$\begin{aligned} f_{anisotropo+umbral}(d_i, a_i|\beta_0, \beta_1, \lambda, \epsilon, \omega) = \ & \beta_0 + \beta_1 \left(1 - \frac{d_i}{u(a_i|\epsilon, \omega)}\right)^+ \\ u(a_i|\lambda, \epsilon, \omega) = & \frac{\lambda}{1 - \epsilon cos(a_i - \omega)} \end{aligned} \tag{A.5}$$

La notación e interpretación de los parámetros y las variables es la misma que en los casos anteriores.

A.2. Implementación

En el Capítulo 3 se han propuesto varias parametrizaciones que permiten modelizar la relación espacial del riesgo en el entorno de un foco contaminante. Las funciones f_{radial} y $f_{anisotropo}$ son lineales en los parámetros y su estimación se realiza mediante los procedimientos estándar para modelos lineales generalizados. Por otra parte, las funciones que incluyen un umbral de alcance ($f_{radial+umbral}$ y $f_{annisotropo+umbral}$) presentan peculiaridades (no linealidad, restricciones en el espacio de los parámetros, discontinuidad en las derivadas, etc) que dificultan la estimación. En este apartado se van a describir técnicas que permitan superar estos inconvenientes en la estimación de los modelos con umbral.

Estimación de los modelos con umbral

La funciones con umbral propuestas (radial con umbral y anisótropa con umbral) están definidas por partes, tienen restricciones en el espacio de los parámetros y éstos presentan relaciones no lineales entre ellos. La metodología estándar de regresión no puede abordar su estimación. A continuación se describe una serie de pasos encaminados a facilitar su estimación. Dado que el modelo radial con umbral es un caso particular del anisótropo con umbral, la presentación se basa en este último. El primero se recupera obviando los parámetros ϵ y ω.

Reparametrización

Algunos parámetros tienen que estar restringidos a un rango para que la función anisótropa represente la realidad que se pretende modelizar.

- La excentricidad de una elipse es, por definición, una cantidad positiva menor o igual que la unidad $\epsilon \in [0, 1]$
- No tiene sentido que el rango de la asociación espacial presente valores negativos ni mayores que la distancia que define el área de estudio. Si se escala la distancia dividiéndola por esa medida (distancia máxima del área de estudio), la restricción sobre λ implica valores positivos menores o iguales a la unidad $\lambda \in [0, 1]$
- El ángulo que forma el eje mayor de la elipse con respecto a la horizontal (o lo que es lo mismo, la dirección de máximo alcance de la

asociación espacial) es una magnitud angular y, como tal, debe restringirse a los primeros 360º (en radianes $\omega \in [0, 2\pi]$)

Una transformación logística proyecta una variable restringida en el intervalo $[0, 1]$ en toda la recta real. Teniendo esto en cuenta, se puede reparametrizar de la siguiente manera

$$\epsilon' = \log \frac{\epsilon}{1-\epsilon};\ \lambda' = \log \frac{\lambda}{1-\lambda};\ \omega' = \log \frac{\omega}{2\pi - \omega} \tag{A.6}$$

y evitar así las restricciones en el espacio de los parámetros al estimarlos en el modelo.

Optimización

Eliminadas las restricciones en el espacio paramétrico, la verosimilitud de los modelos con umbral es una función no lineal y con derivada discontinua, que está determinada de manera unívoca por los datos y la configuración paramétrica.

$$\begin{aligned} &l(\delta^j, \beta_0, \beta_1, \lambda, \epsilon, \omega | O_i, E_i, {z^j}_i, d_i, a_i) = \\ &\sum_{i=1}^{N} O_i \left(log(E_i) + \sum_{j=1}^{M} \delta^j {z^j}_i + \beta_0 + \beta_1 \left(1 - \frac{d_i}{u(a_i|\epsilon,\omega)} \right)^+ \right) - \\ &\sum_{i=1}^{N} E_i e^{\sum_{j=1}^{M} \delta^j {z^j}_i + \beta_0 + \beta_1 \left(1 - \frac{d_i}{u(a_i|\epsilon,\omega)}\right)^+} \end{aligned}$$

Se puede hacer uso de un algoritmo general de optimización para encontrar los valores de los parámetros que maximicen la verosimilitud. Estos valores constituyen la estimación máximo-verosímil dados los datos y las asunciones de distribución.

Una vez en este punto, se dispone de una estimación puntual para cada uno de los parámetros. Quedan por resolver dos cuestiones de inferencia importantes: la estimación de la variabilidad de los parámetros y la comparación entre modelos (existencia de asociación espacial).

Variabilidad de los parámetros

Debido a las características de $f_{radial+umbral}$ y $f_{anisotropo+umbral}$ la estimación usual de la matriz de varianzas-covarianzas de los parámetros, basada en la inversa de la matriz hessiana, puede no resultar adecuada (Seber and Wild 1989; Diggle et al. 1997). Por ello se proponen dos métodos alternativos: uno basado en la verosimilitud y otro en simulaciones de Monte Carlo.

Una manera de estimar la variabilidad de los parámetros es mediante la verosimilitud. Para cada uno de los parámetros $\mu\{\beta_1, \epsilon, \lambda, \omega\}$ y su estimación máximo verosímil μ_{MV}, el intervalo de confianza se define a partir de los valores extremos de μ que son compatibles con la hipótesis $H_0 : \mu = \mu_{MV}$ en un test de razón de verosimilitudes (Venables and Ripley 2002).

Desde un punto de vista práctico, se produce el perfil del logaritmo de la verosimilitud $l(\mu)$ variando el parámetro alrededor de la estimación máximo verosímil μ_{MV}. Después se calcula el valor límite de $l(\mu)$ compatible

con el percentil correspondiente (al nivel de significatividad deseado) de la distribución χ^2 con 1 grado de libertad. Los valores del parámetro para los que el perfil alcanza el límite calculado componen el intervalo de confianza.

Otra manera de estimar la variabilidad de los parámetros es mediante simulaciones de Monte Carlo. Partiendo de los valores de la estimación máximo verosímil y basándose en el modelo, se simulan N conjuntos de datos. Se estima el modelo para cada uno de estos conjuntos de datos simulados. El intervalo de confianza para cada parámetro se obtiene a partir de los percentiles correspondientes de estas N estimaciones. Se trata del mismo procedimiento propuesto por Diggle et al. (1997).

Como se muestra en el Apéndice B, ambos procedimientos de estimación de la variabilidad presentan resultados satisfactorios y similares. Dado que los requerimientos computacionales de la estimación basada en el perfil de verosimilitudes es menor, este procedimiento es preferible.

Comparación de modelos

La parametrización anidada de los modelos permite recuperar el modelo radial ($\epsilon = 0$) o incluso el modelo sin componente espacial ($\beta_1 = 0$). Para contrastar la existencia de asociación espacial y decidir si esta es radial o anisótropa, se puede recurrir a un test de razón de verosimilitudes.

Esta prueba se puede resolver con métodos asintóticos (distribución χ^2) o mediante procedimientos bootstrap (Efron and Tibshirani 1994), que consisten en simular N conjuntos de datos permutando la variable respuesta

(y el offset) y ajustar los dos modelos a comparar. Con esto se obtiene la distribución empírica del estadístico bajo la hipótesis nula, contra la que contrastar el valor del test obtenido con los datos sin permutar.

De nuevo, los resultados del estudio de simulación (Apéndice B) permiten comparar los dos métodos. En este caso, visto el rendimiento de los métodos asintóticos, no está justificado el coste computacional adicional necesario para realizar bootstrap.

APÉNDICE B

Estudio de simulación

Una vez definidos los modelos propuestos y los procedimientos para estimarlos, el siguiente paso es comprobar su comportamiento en situaciones controladas. En esto se basa el estudio de simulación: se generan unos datos "falsos" que responden a una distribución conocida y se aplica sobre ellos cada una de las propuestas para comparar su rendimiento.

Por una parte, interesa conocer cual es la mejor manera de estimar la variabilidad de los parámetros de los modelos con umbral, que está relacionada con la resolución de las pruebas de existencia de asociación espacial (radial y/o anisótropa) con alcance finito. Esto se estudia en la sección B.1.

Por otra, es necesario estudiar cómo afecta la distribución espacial del riesgo a las comparaciones entre las distintas modelizaciones, objetivo de la sección B.2.

Los datos se simularán a partir del modelo más general, es decir el anisótropo, con la configuración paramétrica pertinente en cada caso. En ade-

lante, y para simplificar, se excluyen todas las covariables.

Para llevar a cabo la computación intensiva que estas simulaciones requieren se han paralelizado los procesos con la ayuda de la librería Multicore (Urbanek 2009). Todas las figuras de este apartado se han generado haciendo uso del paquete ggplot2 (Wickham 2009).

B.1. Variabilidad de los parámetros en los modelos con umbral

Planteamiento

Como ya se ha mencionado, los modelos que incorporan un umbral en el alcance de la asociación son modelos no lineales con restricciones en el espacio de los parámetros. Para realizar su estimación se reparametrizan aquellos sujetos a restricciones (λ, ϵ y ω) y se busca la solución que maximice la verosimilitud. Esta metodología es válida para las estimaciones puntuales, pero presenta problemas a la hora de estimar la variabilidad (intervalos de confianza y pruebas de existencia de las diferentes componentes espaciales). Por eso se han propuesto diversas alternativas. En esta sección se compara, mediante simulaciones, las cualidades de cada una de ellas en tres escenarios diferentes:

- Escenario nulo : Sin relación espacial ($\delta = \epsilon = 0$)
- Escenario radial : Sin relación direccional ($\epsilon = 0$)

- Escenario anisótropo: Con asociación espacial radial y direccional ($\epsilon = 0{,}5$)

Para los tres escenarios se asume el mismo número de áreas dentro de la zona de estudio ($N = 200$) y un alcance medio ($\lambda = 0{,}5$). El riesgo relativo en los escenarios en los que sí hay una asociación espacial es de 2. En cada uno de ellos se evalúa la potencia del test de razón de verosimilitudes, que contrasta la existencia de las distintas componentes espaciales, y la cobertura de los intervalos de confianza para los parámetros. Se establece el nivel de significatividad en el 95 % y se simulan 500 conjuntos de datos para cada situación.

Resultados

Dados un contraste que compare dos modelos anidados y un nivel de significatividad, se define la potencia como la probabilidad de rechazar la hipótesis nula (el modelo más sencillo) cuando ésta no es cierta. En un estudio de simulación esta magnitud se puede aproximar por la tasa de rechazos de la hipótesis nula y eso es lo que se presenta a continuación. Para aligerar el seguimiento de la exposición, en adelante se abusa de nomenclatura y a la tasa que aproxima la potencia se le denomina simplemente "potencia".

Conocido el valor real de un parámetro, la cobertura de un intervalo de confianza (a un nivel de significatividad dado) es la probabilidad de que el valor real se encuentre dentro del intervalo de confianza estimado. Dado que en un estudio de simulación se fijan de antemano los valores de

los parámetros, se puede realizar una estimación de la cobertura de los intervalos de confianza para dichos parámetros calculando la proporción de ocasiones en las que el intervalo estimado incluye el valor real. Como en el caso de la potencia, ésta es la magnitud que se presenta a continuación y se denomina "cobertura" a la proporción que estima la cobertura.

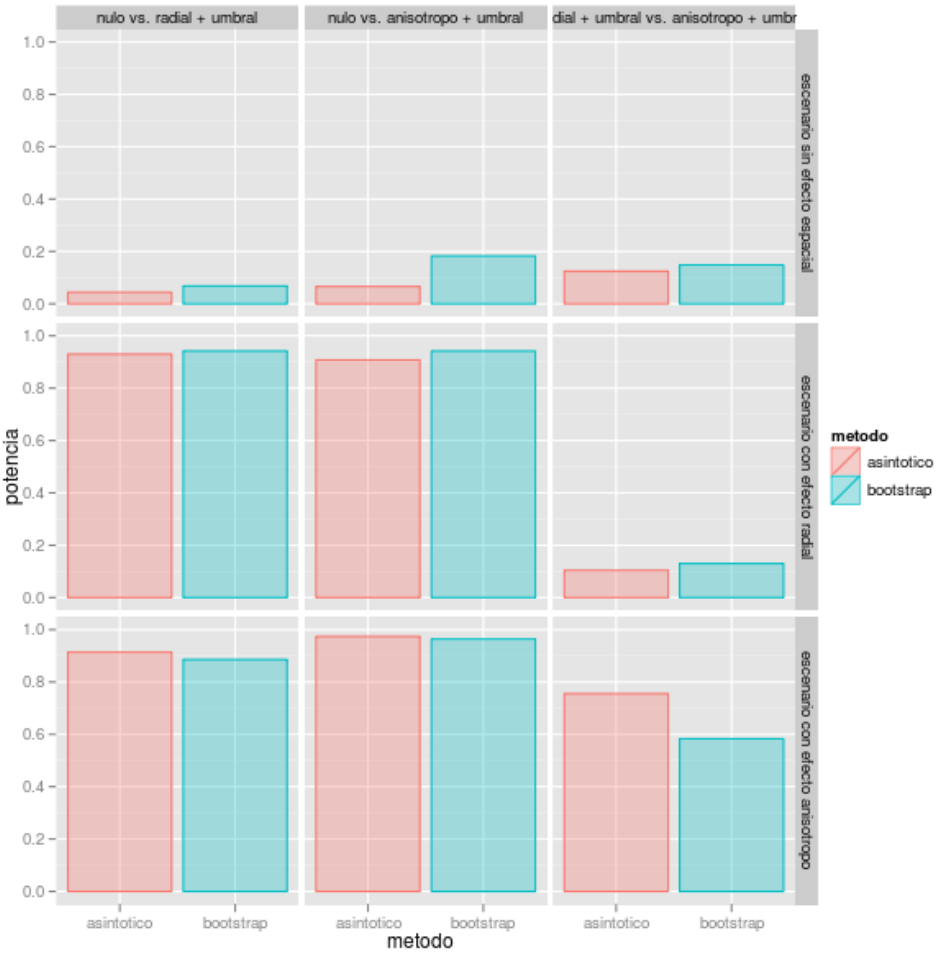


Figura B.1: Comparación de los métodos para los contrastes de existencia de asociación para los modelos con umbral de alcance

Los resultados de la simulación para los contrastes de existencia de asociación espacial, presentados en la Figura B.1, a destacar son los siguientes:

Por una parte, la potencia en el escenario en el que no existe relación espacial (primera fila) es muy baja. Si se tiene en cuenta que se trata de la tasa de falsos positivos, este es un resultado satisfactorio. Por otra, los valores del contraste radial vs. anisótropo en el escenario radial (segunda fila, tercera columna) son los deseables, pues indican que el test rechaza la parametrización anisótropa en situaciones en las que no existe componente direccional en la asociación espacial.

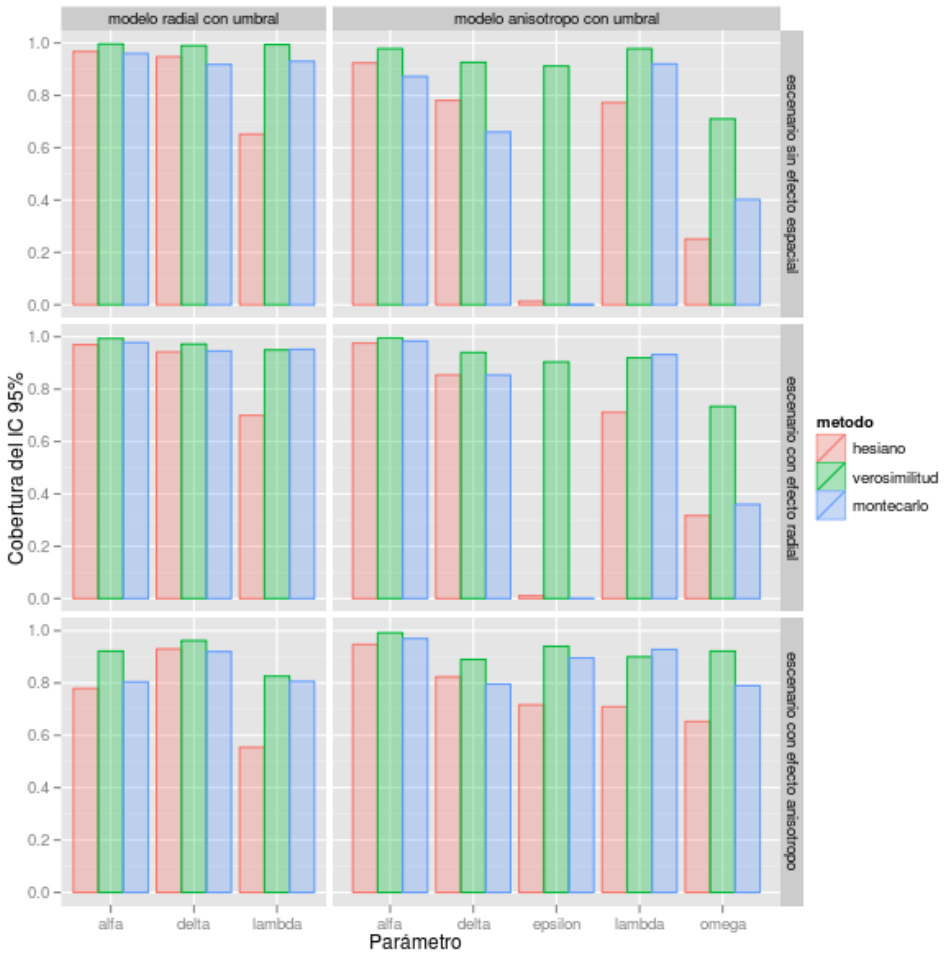


Figura B.2: Comparación de los métodos para la estimación de la variabilidad de los parámetros de los modelos con umbral de alcance en la asociación

En cuanto a los resultados de las simulaciones para comparar los métodos de estimación de la variabilidad de los parámetros, presentados en la Figura B.2, cabe indicar: En primer lugar, la variabilidad estimada mediante el hessiano no resulta fiable. Esto no sorprende, ya que se trata de modelos no lineales. En cuanto a los dos otros métodos, funcionan de manera similar, acercándose a la cobertura nominal.

Conclusiones

Los dos métodos propuestos para resolver la selección de modelos resultan muy parecidos en cuanto a potencia y proporción de falsos positivos. El método asintótico presenta mejor potencia a la hora de detectar componentes direccionales en la asociación espacial cuando éstos sí existen. Esto, unido a que el bootstrap es computacionalmente más costoso, invita a utilizar el primero.

Además de corroborar que el hessiano no es una opción a la hora de estimar la variabilidad de los parámetros en estos modelos, se puede concluir que la variabilidad se estima razonablemente bien por cualquiera de los dos otros métodos. Al igual que pasaba con el bootstrap en los contrastes, el método de Monte Carlo consume mayores recursos computacionales y temporales. Por ello, es preferible estimar la variabilidad mediante el perfil de verosimilitudes.

Por lo tanto, y si no se indica lo contrario, la selección de modelos se realizará mediante contrastes de razón de verosimilitudes asintóticos y la

estimación de la variabilidad de los parámetros mediante el perfil de verosimilitudes.

B.2. Comparación de modelos

Planteamiento

El objetivo de esta sección es determinar en qué medida la distribución del riesgo real (y conocida, pues se trata de un estudio de simulación) condiciona la capacidad de las distintas propuestas de detectar asociaciones espaciales. Para ello se simulan 10.000 conjuntos de datos en los que el valor de los parámetros relevantes se establece de manera aleatoria entre un rango plausible:

- Dirección de máximo riesgo $\omega \in [0, 2\pi]$
- Excentricidad $\epsilon \in [0, 1]$
- Alcance $\lambda \in [0, 50]$ Km
- Riesgo en el foco $RR \in [1, 3]$
- Número de municipios $N \in [50, 200]$
- Tasa de mortalidad $t \in [15, 45]$ defunciones por 100.000 habitantes

En cada uno de los escenario simulados se ajustan los modelos (Nulo, Radial, Radial con Umbral, Anisótropo y Anisótropo con Umbral) y se realizan las comparaciones relevantes dos a dos:

- Radial vs. Nulo

- Radial con Umbral vs. Radial
- Anisótropo vs. Radial
- Anisótropo con Umbral vs. Radial con Umbral
- Anisótropo con Umbral vs. Anisótropo

Las comparaciones se cuantifican mediante (el logaritmo de) la razón de verosimilitudes. Este estadístico se relaciona de manera directa con los odds del modelo "complejo" frente al "sencillo". Refleja la capacidad de detectar la componente espacial considerada en cada comparación.

Resultados

Como se dispone de un valor para cada una de las comparaciones en cada uno de las situaciones simuladas (5 x 10.000), es necesario resumir esta información de manera que sea interpretable. Para ello se realiza una suavización mediante splines cúbicas de las seis variables que intervienen para cada una de las 5 comparaciones. Los resultados se presentan en forma de gráficos de lineas con bandas de confianza al 95 %, separados para cada parámetro: El logaritmo de la razón de verosimilitudes en el eje Y, el valor de cada parámetro en el X. Cada color corresponde a una de las comparaciones (Figura B.3).

En cada gráfica se presenta la variación de uno de los parámetros. El resto se mantiene constante en su valor medio (de entre los valores de la simulación). Esto condiciona el valor absoluto de la razón de verosimilitudes, pero no su valor relativo. Por ello, la información relevante de las gráficas es

la forma de las curvas suavizadas y las posiciones relativas entre ellas, pero no la escala absoluta.

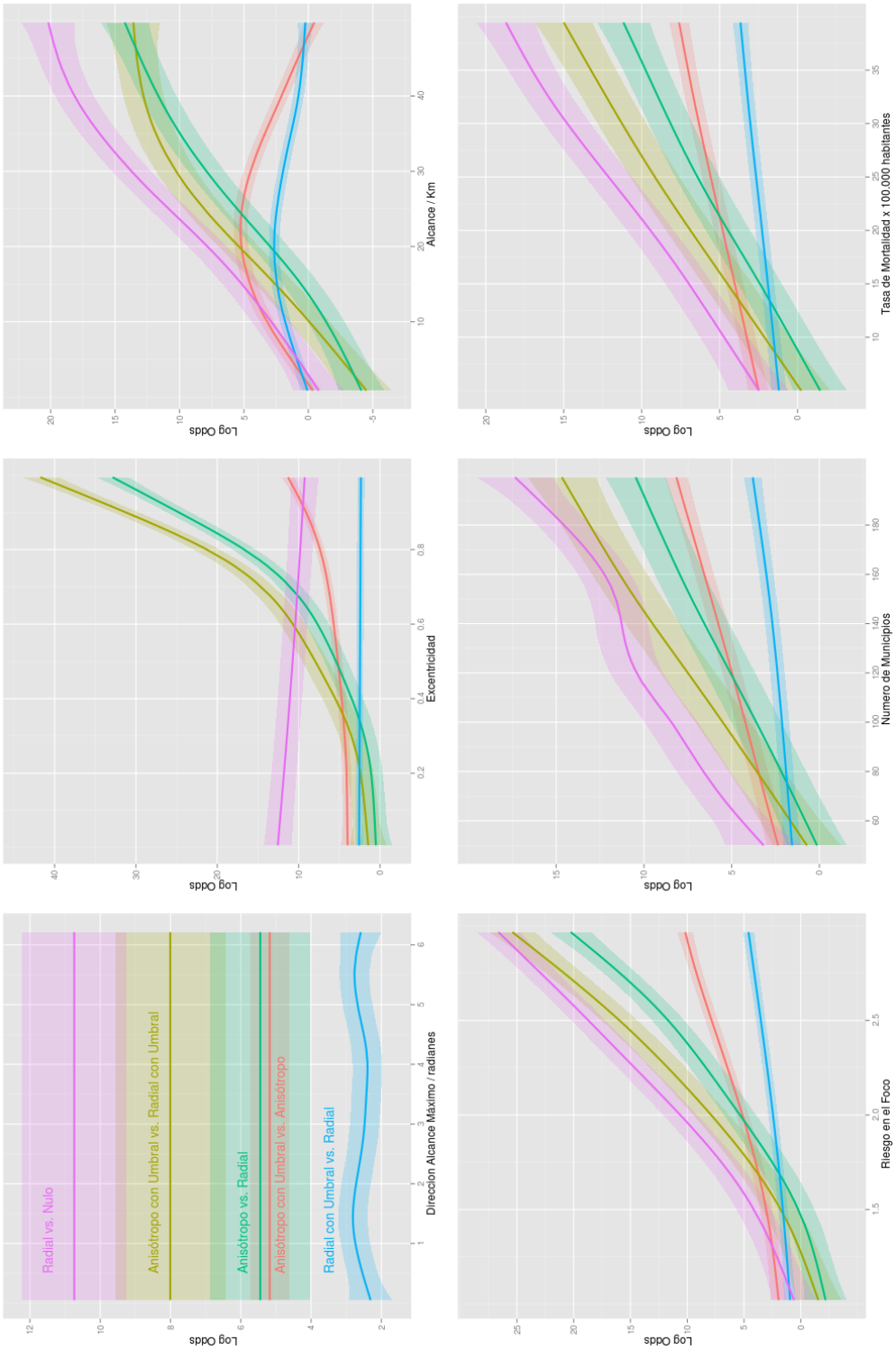


Figura B.3: Variación de las comparaciones entre modelos con las características de los escenarios

En el primer gráfico de la Figura B.3 se observa que el resultado de la comparación entre cualquiera de las parejas de modelos no depende de la dirección del riesgo. Este resultado es positivo, ya que lo contrario indicaría problemas en el proceso de estimación.

Con respecto a la excentricidad, se observan 3 tipos de patrones. Por una parte, el rendimiento de los modelos que no tienen en cuenta la dirección decrece al aumentar la excentricidad. Por otra, las comparaciones en las que se contrasta la existencia de una componente direccional se ven fuertemente favorecidas por excentricidades grandes. Por último, la mejora que supone incluir un umbral en el modelo anisótropo también mejora al aumentar la excentricidad, pero de manera menos pronunciada.

Si aumenta el alcance de la asociación, cuando no se está contrastando directamente la existencia de un umbral, crecen los odds en favor del modelo complejo. Sin embargo, al comparar directamente la existencia de un umbral, el mejor rendimiento se encuentra para alcances medios.

El riesgo en el foco, número de municipios y tasa de mortalidad, variables todas relacionadas con el tamaño muestral y la intensidad de la asociación, presentan proporcionalidad directa con el rendimiento de los contrastes.

Conclusiones

En los escenarios en los que existe una distribución direccional del riesgo y/o una asociación espacial con alcance efectivo no lejano al foco, los métodos propuestos son capaces de mejorar sustancialmente la detección

de asociaciones espaciales. Los modelos radial con umbral y anisótropo sin umbral permiten descripciones intermedias.

En particular, con respecto a la direccionalidad, los modelos anisótropos obtienen resultados satisfactorios para excentricidades moderadas y altas.

Para entender el efecto del alcance, hay que tener en cuenta que éste actúa como un limitador del tamaño muestral; si el alcance es corto, existen pocos municipios expuestos y, por tanto, resulta difícil detectar esa asociación. Por contra, si es largo, prácticamente toda la zona estudiada resulta expuesta, con lo que introducir un umbral no mejora la descripción (los modelos sin umbral se comportan de manera parecida a los que incorporan umbral). Por eso las situaciones en las que resulta ventajosa la parametrización con umbral son aquellas en las que el alcance es medio.