

CONCATENACIÓN TEMPORAL DE MODELOS ESPACIALES Y SU APLICACIÓN AL ESTUDIO DE LA MENINGITIS EN ESPAÑA

JUAN FERRÁNDIZ FERRAGUD*
FERRAN MARTÍNEZ I NAVARRO**
PILAR SANMARTÍN FITA*

La cartografía de enfermedades infecciosas en periodos sucesivos plantea la necesidad de su extensión al caso dinámico. En este trabajo proponemos la concatenación temporal de modelos auto-regresivos espaciales para abordar el análisis de mortalidad por meningitis en España en el período 1950 – 1990 con datos agregados a nivel provincial. Para la estimación y selección del modelo usamos técnicas basadas en la función de verosimilitud

Temporal concatenation of spatial models and its application to the study of meningitis in Spain

Palabras clave: Análisis espacio-temporal, modelos espaciales auto-regresivos, enfermedades infecciosas

Clasificación AMS (MSC 2000): 62M10, 62M30, 62P10

* Departament d'Estadística i I.O., Universitat de València, València.

** Instituto de Salud Carlos III, Madrid.

* Departamento de Matemática Aplicada y Estadística, Universidad Politécnica de Cartagena, Paseo Alfonso XIII s/n, 30203 Cartagena.

– Recibido en junio de 1999.

– Aceptado en enero de 2001.

1. INTRODUCCIÓN

En epidemiología, los estudios geográficos analizan datos correspondientes a distintas localizaciones espaciales a lo largo de un determinado período temporal. En este contexto surge la necesidad de un análisis espacial a la vez que temporal para entender la evolución de las dependencias espaciales a lo largo del tiempo. El éxito alcanzado con modelos jerárquicos en la cartografía de enfermedades gracias a la disponibilidad de métodos MCMC, ha potenciado su extensión al análisis espacio-temporal de datos geográficos de morbilidad/mortalidad [véase por ejemplo Waller *et al.* (1997); Knorr-Held y Besag (1998)].

Estos modelos recogen la dependencia mutua espacial de las observaciones indirectamente, a través de factores ocultos en el segundo nivel de la jerarquía. Este modo de expresar la dependencia reflejaría el hecho de que el riesgo en una localización depende de factores de tipo medio-ambiental que seguirían un determinado patrón geográfico sobre el área de estudio.

Sin embargo, cuando tratamos con enfermedades infecciosas, cabe esperar la influencia directa de la mortalidad/morbilidad en regiones próximas. Parece más adecuado recurrir en estos casos a modelos espaciales auto-regresivos (distribuciones de Gibbs) y, en particular, a los auto-modelos de Besag (1974) por su facilidad de interpretación.

Las distribuciones de Gibbs se originaron en el marco de las ciencias físicas a principios de siglo para modelizar sistemas interactivos de partículas [Gibbs (1902)] y han tenido un desarrollo posterior en el marco de otras ciencias. Se incorporan al estudio de análisis de imágenes digitalizadas en la década pasada, siendo básico en este campo el trabajo de Geman y Geman (1984). En el campo de la estadística espacial se introducen en la década de los 70 siendo de gran relevancia la contribución de Besag (1974), que propone una familia de modelos espaciales así como procedimientos para su análisis y contraste.

Más recientemente, el propio Besag adapta las técnicas del análisis de imágenes digitalizadas al análisis de datos epidemiológicos espaciales utilizando procesos de variables aleatorias dispuestas en retículos irregulares. Por otra parte Darroch *et al.* (1980) analizan la similitud de los campos de Gibbs con el estudio de las interacciones entre factores en tablas de contingencia, dentro del contexto de los modelos gráficos.

Sirva como ejemplo de aplicaciones recientes en este contexto los trabajos de Ferrándiz *et al.* (1995) y Ferrándiz *et al.* (1999) que aplican estos modelos al estudio de la mortalidad por cáncer en Valencia y la posible influencia de la contaminación por nitratos del agua potable. Asimismo, en López (1997) se desarrollan este tipo de distribuciones introduciendo los modelos lineales generalizados espaciales.

En el presente trabajo consideramos la extensión de las distribuciones de Gibbs al caso espacio-temporal con el fin de analizar la mortalidad anual por meningitis en España en el período 1950-90 con datos agregados provincialmente.

Asimismo pretendemos estudiar la hipótesis establecida por Martínez (1989) de la relación existente entre reproducción social y enfermedad. El incremento generalizado de las tasas coincide con los cambios estructurales de la sociedad española y concretamente con el proceso de urbanización generado por la forma que asumió la reproducción social en el modelo del desarrollo económico y social de las décadas del 60 y 70 (Plan de Estabilización (1959) y planes de Desarrollo (1964-1979)) orientando las migraciones de las zonas rurales a los polos de desarrollo. Los movimientos migratorios influirían en la aparición de los brotes epidémicos, alterando considerablemente la regularidad natural de las sucesivas ondas epidémicas.

La inmigración constituye un incremento de la población susceptible a la enfermedad que interfiere con la pauta natural impuesta por el crecimiento vegetativo en cada provincia y provoca un cambio en el ciclo natural de la inmunización de la población. Los datos considerados corresponden a las 50 provincias españolas, excluyendo Ceuta y Melilla por su heterogéneo tratamiento en las estadísticas oficiales a lo largo del período. Junto con las tasas de mortalidad, se han recogido los datos de inmigración anual provincial.

En la sección 2 presentamos el modelo a utilizar, que se basa en una concatenación temporal de distribuciones espaciales auto-regresivas. En la sección 3 estudiamos el ajuste del modelo y en las secciones 4 y 5 presentamos, respectivamente, su aplicación a la mortalidad por meningitis y las conclusiones obtenidas.

2. MODELO

Las observaciones constituyen una serie temporal multivariante $\{\mathbf{y}(t)\}_{t=1}^T$ de manera que $\mathbf{y}(t) \in \mathbb{R}^N$ es un vector que recoge los valores observados en N localizaciones espaciales en el instante t . Dada la estructura espacial de los datos en cada instante temporal, abordaremos el problema como una concatenación temporal de distribuciones espaciales. Consideraremos distribuciones de Gibbs [ver Cressie (1991)].

2.1. Distribuciones espaciales

Las distribuciones de Gibbs (equivalentemente campos aleatorios Markovianos) definen la distribución conjunta a partir de las distribuciones condicionadas con densidad de probabilidad:

$$(1) \quad p(y_i | \mathbf{y}_{-i})$$

siendo \mathbf{y}_{-i} el subvector que se obtiene al excluir la componente i -ésima del vector \mathbf{y} . Asimismo la densidad (1) involucra únicamente a un subconjunto δ_i de localizaciones que se denominan «vecinas» de la localización i de forma que la anterior expresión coincide con:

$$p(y_i | \{y_j : j \in \delta_i\})$$

Los auto-modelos de Besag [Besag (1974)] son un caso particular de distribuciones de Gibbs caracterizados por el hecho de que las distribuciones condicionadas (1) pertenecen a la familia exponencial y el efecto de las regiones vecinas se incorpora a través de términos que involucran únicamente pares de localizaciones:

$$(2) \quad p(y_i | \mathbf{y}_{-i}) \propto \exp\left\{ \left(a_i + \sum_{j \in \delta_i} \beta_{ij} y_j \right) y_i + h_i(y_i) \right\}.$$

En la expresión (2) β_{ij} son parámetros de interacción espacial y a_i es un parámetro específico de cada localización, en el cual recogeremos la posible influencia de un vector de covariables $\mathbf{x}_i \in \mathbb{R}^R$ observadas mediante:

$$(3) \quad a_i = \mathbf{x}_i' \boldsymbol{\alpha}_i$$

con lo cual tendríamos un modelo lineal generalizado espacial [Ferrándiz *et al.*, (1995); López (1997)]. Cuando usamos la distribución Gaussiana para modelizar las distribuciones condicionadas tenemos el caso particular del modelo CARX, auto-Gaussiano condicionalmente especificado con covariables [ver Cressie (1991); Richardson *et al.* (1992)]:

$$(4) \quad y_i | \mathbf{y}_{-i} \sim N\left(\mathbf{x}_i' \boldsymbol{\alpha}_i + \sum_{j \in \delta_i} \beta_{ij} (y_j - \mathbf{x}_j' \boldsymbol{\alpha}_j), \sigma_i^2 \right)$$

de donde se deduce la distribución conjunta:

$$(5) \quad \mathbf{y} \sim N(X\boldsymbol{\alpha}, (I - B)^{-1} \text{Diag}(\sigma_i^2))$$

La matriz B tiene como elementos los coeficientes β_{ij} , y ceros en su diagonal. $\text{Diag}(\sigma_i^2)$ es una matriz diagonal con elementos σ_i^2 . Las columnas de la matriz X corresponden a las covariables incluidas en el modelo, y $\boldsymbol{\alpha} = (\boldsymbol{\alpha}'_1, \dots, \boldsymbol{\alpha}'_N)'$.

2.2. Modelos espaciales dinámicos

Partiendo de los anteriores modelos planteamos su concatenación temporal de la siguiente forma:

$$(6) \quad p(y_i(t) | \mathbf{y}(t)_{-i}, \mathbf{y}(t)^*, \mathbf{x}_i(t), \mathbf{x}_i(t)^*) \propto \exp\left\{a_i(t) + \sum_{j \in \delta_i} \beta_{ij}(t) y_j(t) y_i(t) + h_i(y_i(t))\right\}.$$

En la expresión (1) los vectores $\mathbf{x}_i(t)^*$ e $\mathbf{y}(t)^*$ representan el pasado de $\mathbf{x}_i(t)$ e $\mathbf{y}(t)$ respectivamente. La influencia de las covariables que constituyen \mathbf{x}_i y su pasado \mathbf{x}_i^* , así como la influencia del propio pasado \mathbf{y}^* , se ejerce a través del parámetro

$$a_i(t) = a_i(\mathbf{y}(t)^*, \mathbf{x}_i(t), \mathbf{x}_i(t)^*)$$

dando lugar a los auto-modelos dinámicos [ver Sanmartín (1997)].

En particular, para el caso Gaussiano la expresión anterior resulta ser,

$$(7) \quad \mathbf{y}(t) | X(t), \mathbf{y}(t)^*, X(t)^* \sim N(\boldsymbol{\mu}(t), (I - B(t))^{-1} \text{Diag}(\sigma_i^2(t)))$$

de modo que el pasado y las covariables ejercen su influencia a través del vector de medias

$$\boldsymbol{\mu}(t) = \boldsymbol{\mu}(X(t), \mathbf{y}^*(t), X(t)^*)$$

Atendiendo a criterios de proximidad geográfica y temporal, y buscando su expresión en forma de dependencia lineal, el vector de medias $\boldsymbol{\mu}(t)$ se ha especificado de la forma:

$$(8) \quad \boldsymbol{\mu}(t) = \sum_{i=1}^p A_{0i} \mathbf{y}(t-i) + \sum_{r=1}^R \sum_{j=0}^{q_r} A_{rj} \mathbf{x}_r(t-j)$$

donde las matrices A_{0i} y A_{rj} son matrices de dependencia espacial y cada covariable puede actuar con un número diferente de retardos temporales. En la expresión (8) hemos permitido que las covariables en las localizaciones «vecinas» y su pasado también influyan en la localización i -ésima, relajando la restricción de su influencia exclusivamente local expresada en (1). Esta posible simplificación queda recogida sin más que definir las A_{rj} como matrices diagonales o incluso proporcionales a la identidad y $q_r = 1$.

Estas matrices de dependencia espacial se construyen habitualmente como combinaciones lineales de matrices fijas [ver Cressie (1991); Stoffer (1986)]. En nuestro caso hemos establecido las matrices A_{0i} y A_{rj} a partir de matrices de proximidades geográficas que reflejan el posible intercambio de gente entre cada dos provincias y recogen una «gradación» en el orden de vecindad de las localizaciones (en la sección 4 se desarrollan con más detalle).

La misma estructura de proximidades se ha utilizado para la matriz $B(t)$ de coeficientes auto-regresivos espaciales, aunque anulando sus elementos diagonales y tomándola

simétrica como es preceptivo en el modelo CARX. Hemos considerado $B(t) \equiv B$, fija a lo largo del tiempo, es decir,

$$(9) \quad \begin{aligned} A_{0i} &= \sum_{k=1}^m \beta_{oik} D_k & i = 1, \dots, p \\ A_{rj} &= \sum_{k=1}^m \beta_{rjk} D_k & r = 1, \dots, R \quad j = 0, \dots, q_r \\ B &= \kappa(D - \text{Diag}(D)) \end{aligned}$$

con D_k $k = 1, \dots, m$ y D matrices fijas.

Los parámetros β miden el amortiguamiento del efecto con el tiempo, y el parámetro κ debe cumplir:

$$|\kappa| < \min_{\lambda_i} (1/|\lambda_i|)$$

donde $\{\lambda_i\}$ es el conjunto de valores propios de la matriz $(D - \text{Diag}(D))$ a fin de asegurar que la matriz de varianzas y covarianzas sea definida positiva [ver Cressie (1991)]. También hemos asumido la simplificación

$$\text{Diag}(\sigma_i^2(t)) \equiv \sigma^2 I_N$$

Dado que las observaciones son datos de mortalidad parecería más adecuado abordar el problema usando la distribución auto-Poisson, modelo que se aplica suponiendo que la variable observada es el número de casos. Sin embargo, en muchas aplicaciones médicas se requiere la modelización directa de las tasas (como puede verse en los ejemplos de los párrafos siguientes), y es ésta la situación que nosotros pretendemos abordar.

Por otra parte, los modelos auto-regresivos de Poisson con covariables conllevan un grado de complejidad elevado [como ejemplo de aplicación para el caso estático ver Ferrándiz et al. (1995); Ferrándiz *et al.* (1999)]. El uso de modelos auto-regresivos Gaussianos presenta un menor grado de complejidad y tiene una larga tradición en el análisis de datos de mortalidad y morbilidad en epidemiología espacial que aún sigue vigente.

Cliff *et al.* (1975) y Cliff *et al.* (1992), asumiendo datos Gaussianos, recurren a modelos ARMA espacio-temporales para analizar el sarampión en Islandia. Martin y Oeppen (1975) aplican estos modelos al estudio de la peste animal. Cook y Pocock (1983) modelizan las tasas estandarizadas de mortalidad cardiovascular mediante el uso de un modelo auto-Gaussiano espacial. Cliff y Hagget (1993) hace un resumen de las técnicas estadísticas para el análisis de datos relacionados con las epidemias de gripe y sarampión. En él se recoge la aplicación de los modelos Box-Jenkins al caso espacio-temporal (modelos STARMA) para modelizar el número de casos registrados. Más

recientemente, en el libro editado por Fisher y Getis (1997) que recoge las últimas técnicas de análisis espacial, aparece el trabajo de Logan y Cliff (1997) aplicando un modelo estructural Gaussiano para predecir la incidencia de una epidemia.

El presente trabajo se adhiere a esta corriente incorporando al modelo la dependencia espacial, simultánea y diferida, con el fin de comprobar el efecto de regiones próximas.

3. AJUSTE DEL MODELO

Sea $s = \max\{p, q_1, \dots, q_r\}$ el máximo retardo temporal en la expresión (8). A partir de la definición (7), si condicionamos el análisis estadístico a los s instantes iniciales, la función de log-verosimilitud se puede construir con el producto de las sucesivas distribuciones condicionadas

$$(10) \quad \begin{aligned} \ell(\beta, \kappa, \sigma^2) &= \log\left(\prod_{t=s+1}^T p(\mathbf{y}(t) | X(t), \mathbf{y}(t)^*, X(t)^*)\right) \\ &= -\frac{1}{2\sigma^2} Q(\kappa, \beta) + \frac{T-s}{2} \log |I - B| + \\ &\quad - \frac{(T-s)N}{2} \log \sigma^2 + \text{const.} \end{aligned}$$

siendo

$$(11) \quad Q(\kappa, \beta) = \sum_{t=s+1}^T (\mathbf{y}(t) - \boldsymbol{\mu}(t))' (I - B) (\mathbf{y}(t) - \boldsymbol{\mu}(t)).$$

La estimación máximo-verosímil puede obtenerse expresando (10) como log-verosimilitud *perfil* de κ extendiendo al caso dinámico el procedimiento propuesto para el caso estático en Cressie (1991) y Richardson *et al.* (1992):

- para κ fija, resolviendo el problema de mínimos cuadrados generalizados

$$(12) \quad \hat{\beta} = \underset{\beta}{\operatorname{argmin}} Q(\kappa, \beta)$$

dando lugar a la suma de cuadrados residual

$$Q^*(\kappa) = Q(\kappa, \hat{\beta})$$

- para el mismo valor fijo de κ , optimizando respecto de σ^2

$$\hat{\sigma}^2 = \frac{1}{(T-s)N} Q^*(\kappa)$$

- finalmente, maximizando la log-verosimilitud *perfil* de κ

$$(13) \quad \hat{\kappa} = \underset{\kappa}{\operatorname{argmax}} \ell(\kappa, \hat{\beta}(\kappa), \hat{\sigma}^2(\kappa))$$

Con la asignación (10), el problema de mínimos cuadrados generalizados de (12) corresponde al diseño:

$$(14) \quad \mathbf{y} = X\beta + \mathbf{u}$$

con: $\mathbf{y}' = (\mathbf{y}(T)', \dots, \mathbf{y}(s+1)')$,

$\mathbf{u}' = (\mathbf{u}(T)', \dots, \mathbf{u}(s+1)'),$ y

$\beta' = (\beta_{011}, \dots, \beta_{RmqR}).$

Las columnas de la matriz de diseño X se deducen de (9), de modo que la que corresponde al parámetro β_{0jk} es:

$$(15) \quad (I_{T-s} \otimes D_k)(\mathbf{y}(T-j)', \mathbf{y}(T-j-1)', \dots, \mathbf{y}(s-j+1)')$$

siendo I_{T-s} la matriz identidad de orden $T-s$ y el operador \otimes el producto Kronecker de matrices. Para columnas correspondientes a β_{rjk} con $r > 0$, bastará sustituir en (15) los vectores $\mathbf{y}(t)$ por los correspondientes a la r -ésima covariable $\mathbf{x}(t)$.

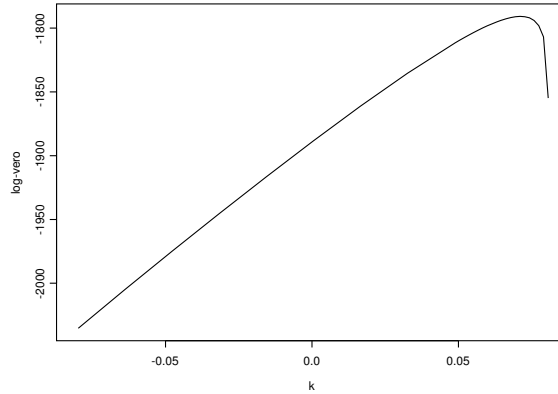


Figura 1. Gráfica de la log-verosimilitud perfil para el modelo 3.

La matriz de varianzas y covarianzas del vector \mathbf{u} resulta ser $I_{T-s} \otimes (I - B)^{-1} \sigma^2$, y es la responsable de que (14) no sea un problema de mínimos cuadrados ordinarios sino generalizados.

La optimización directa de la log-verosimilitud perfil (13) es ahora posible. Para algunos esquemas de parametrización de la matriz de varianzas y covarianzas en el caso puramente espacial esta función puede presentar multimodalidad como Richardson *et al.* (1992) señalan [ver asimismo Mardia y Marshall (1984); Mardia y Watkins (1989); Warnes y Ripley (1987)]. Mardia y Watkins (1989) recomiendan un estudio gráfico de la log-verosimilitud perfil. En nuestro caso (ver figura 1) dicha función es unimodal y cóncava.

Por las propiedades de la log-verosimilitud perfil [ver McCullagh y Nelder (1989)], el máximo de dicha función coincide con el de la log-verosimilitud y se alcanza también en el estimador máximo verosímil.

Para la optimización directa de (13) hemos usado las rutinas de cálculo de mínimos cuadrados generalizados y optimización univariante existentes en el programa S-Plus [ver Venables y Ripley (1999); Becker *et al.* (1988)]. Para esquemas más complicados de la log-verosimilitud perfil se puede recurrir a algún algoritmo numérico como por ejemplo el basado en el *downhill simplex method* de Nelder y Mead [ver Press *et al.* (1986)].

Se puede valorar la incorporación de los distintos términos a la esperanza condicionada $\mu(t)$, así como la hipótesis de independencia espacial simultánea, ($\kappa = 0$), mediante la diferencia de la log-verosimilitud conseguida en cada caso, comparándola con la distribución asintótica χ^2 del cociente de verosimilitudes. Para la obtención de errores estándar aproximados de los estimadores de los parámetros asociados a los distintos términos de la esperanza condicionada hemos hecho uso de las técnicas empleadas en regresión múltiple para un problema de mínimos cuadrados generalizados, en la misma línea sugerida en Pfeifer y Deutsch (1980) para el caso de modelos auto-regresivos espaciales Gaussianos sin interacción espacial simultánea. Este procedimiento también es usado por Richardson *et al.* (1992) para el caso CARX estático [ver asimismo Mardia *et al.* (1984)].

4. MORTALIDAD POR MENINGITIS EN ESPAÑA (1950-1990)

Los datos analizados en el presente trabajo corresponden a tasas de mortalidad anuales para las provincias españolas salvo Ceuta y Melilla, que no han sido incluidas por su heterogéneo tratamiento en las estadísticas oficiales. Junto con las tasas de mortalidad, se han recogido los datos de inmigración anual a lo largo del período.

Como variable respuesta se ha tomado la raíz cuadrada de las tasas para estabilizar la varianza tal y como sugerían los resultados de análisis exploratorios previos. Es similar a la transformación de Freeman-Tuckey utilizada en Cressie (1991), pero no requiere el uso explícito de la mortalidad cruda.

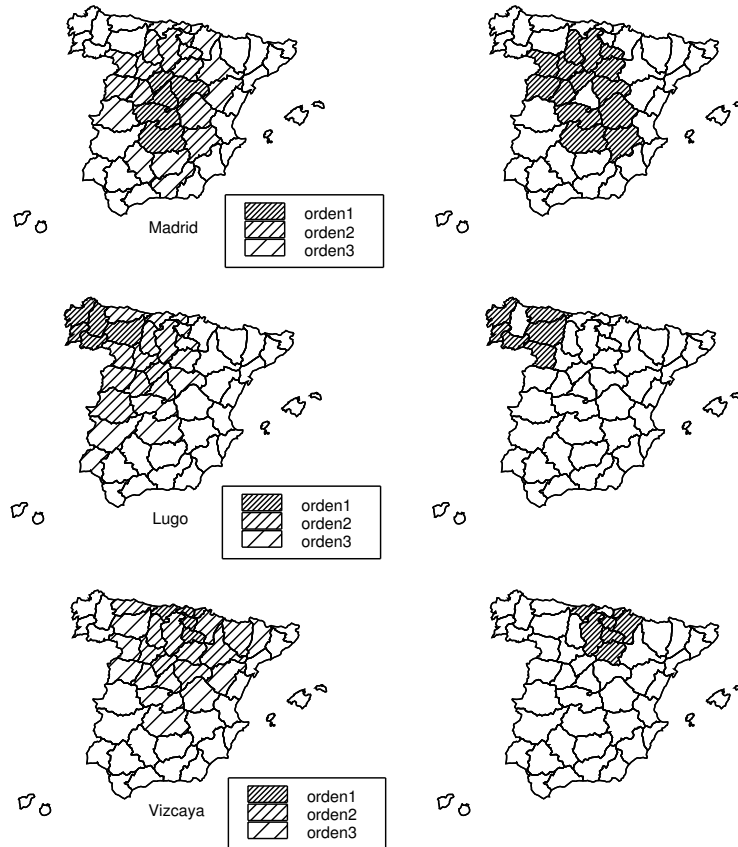


Figura 2. Sistema de vecindades espaciales.

También hemos considerado un retardo temporal de orden $p = 5$ y orden espacial $m = 3$ para el proceso auto-regresivo. Las matrices D_k , $k = 1, 2, 3$ consideradas en (9) están formadas por ceros y unos atendiendo a una medida de proximidad entre las localizaciones. Este índice de proximidad consiste en calcular el inverso de la distancia (entre capitales de provincia). Para cada provincia hemos clasificado sus vecinos según el índice esté comprendido entre determinados cuantiles de las «proximidades» (de ella misma con el resto de provincias), vecinos de orden 1, D_1 , (proximidades superiores al percentil 90), vecinos de orden 2, D_2 (superiores al percentil 70 e inferiores o iguales al 90), vecinos de orden 3, D_3 (superiores al percentil 50 e inferiores o iguales al 70). Y como matriz D hemos considerado la totalidad de «proximidades» interprovinciales y hemos considerado vecinas aquellas provincias cuyo índice supera el percentil 80 (la matriz así obtenida es simétrica).

En la figura 2 aparecen representados los vecinos de algunas provincias (en la primera columna vecinos en retardo temporal (D_1 , D_2 y D_3) y en la segunda columna vecinos en el mismo instante temporal D).

En cuanto a covariables hemos considerado la variable inmigración en cada localización sin considerar la posible influencia de los valores observados en localizaciones vecinas (influencia local, $A_{1,0} = I$), y retardo temporal $q_1 = 1$ (tras observar que considerar un retardo de $q_1 = 5$ como en la parte auto-regresiva no mejoraba significativamente el modelo). Finalmente, como también sugerían los análisis previos, consideramos una tendencia polinomial cúbica diferente para cada provincia.

Con todo ello, los modelos ajustados son los siguientes:

modelo 1 Tendencia temporal polinomial cúbica diferente para cada provincia.

modelo 2 Modelo espacio-temporal auto-regresivo con tendencia temporal local (orden temporal 5, orden espacial 3).

modelo 3 Modelo espacio-temporal auto-regresivo con tendencia temporal local (orden temporal 5, orden espacial 3) con covariable inmigración (orden temporal 1, influencia local).

Un resumen de los resultados aparece en la tabla 1 siendo ℓ el logaritmo de la verosimilitud. Como comentamos en la sección anterior, hemos usado el cociente de verosimilitudes para comparar la significación de cada modelo considerando la correspondiente distribución χ^2 y un nivel de significación de 0.05.

Tabla 1. Comparación de modelos.

modelo	$\kappa = 0$			$\kappa \neq 0$			$\Delta\ell(\text{g.l.})$ $\kappa = 0$ vs $\kappa \neq 0$
	ℓ	SSE	$\Delta\ell$ (g.l.)	ℓ	SSE	$\Delta\ell$ (g.l.)	
1	-2108.47	1097.08		-1895.49	817.61		212.98(1)
2	-1892.60	863.56	215.87 (15)	-1795.78	742.45	99.71 (15)	96.82(1)
3	-1889.27	860.37	3.33 (2)	-1790.86	738.24	4.92 (2)	98.41(1)

Los tres modelos se han estimado suponiendo $\kappa = 0$ que es equivalente a considerar que no hay interacción espacial simultánea y $\kappa \neq 0$. El incremento en la log-verosimilitud

para cada uno de los tres modelos aparece al final de cada fila y si consideramos un nivel de significación de 0.05 podemos aceptar interacción espacial simultánea frente al modelo de no interacción en los tres casos (el percentil de la distribución χ^2 con $n = 1$ grados de libertad es 3.841).

Tabla 2. Estimación de los parámetros del Modelo 3.

Parte				
auto-regr.	coef	err. std.	stat. t.	p.valor
β_{011}	0.0927	0.0112	8.2707	0.0000
β_{012}	0.0051	0.0072	0.7147	0.4749
β_{013}	-0.0062	0.0075	-0.8245	0.4098
β_{021}	-0.0107	0.0124	-0.8645	0.3874
β_{022}	-0.0149	0.0077	-1.9278	0.0541
β_{023}	0.0009	0.0082	0.1150	0.9085
β_{031}	0.0108	0.0125	0.8639	0.3878
β_{032}	0.0000	0.0078	0.0031	0.9975
β_{033}	-0.0238	0.0083	-2.8634	0.0042
β_{041}	-0.0115	0.0125	-0.9191	0.3582
β_{042}	-0.0026	0.0078	-0.3365	0.7366
β_{043}	0.0000	0.0084	0.0057	0.9954
β_{051}	-0.0395	0.0114	-3.4697	0.0005
β_{052}	0.0187	0.0074	2.5113	0.0121
β_{053}	-0.0169	0.0078	-2.1698	0.0302
Inmig.	coef ($\times 10^{-3}$)	err. std. ($\times 10^{-3}$)	stat. t.	p.valor
β_{101}	0.0035	0.0033	1.0464	0.2955
β_{111}	0.0068	0.0032	2.1025	0.0357
Interacción	simultánea			
$\hat{\kappa} = 0.0712$		$\hat{\sigma}^2 = 0.4101$		

Al mismo tiempo, para ambas hipótesis ($\kappa = 0$ o $\kappa \neq 0$) la incorporación de la parte auto-regresiva espacial y temporal (modelo 2) y la introducción de la covariable inmigración (modelo 3) es significativa frente al modelo anterior (se observa que son modelos anidados) a un nivel de significación de 0.05. Los percentiles de la distribución χ^2 con $n = 15$ y $n = 2$ grados de libertad son 24.996 y 5.991 respectivamente, que

son superados en cada caso por el doble del logaritmo del cociente de verosimilitudes, $2 \Delta \ell$).

En la tabla 2 aparecen los valores estimados de los parámetros del Modelo 3, junto con sus errores estándar aproximados. Vemos que son significativamente distintos de cero, a nivel individual y en presencia del resto de covariables, la influencia de los vecinos de orden uno espacial en un retardo temporal de un año, de orden tres espacial en un retardo de tres años y orden espacial uno, dos y tres en un retardo de cinco años, asimismo es «casi» significativamente distinta de cero, la influencia de los vecinos de orden dos espacial en un retardo temporal de dos años. La interacción simultánea viene caracterizada a través de κ y la matriz de vecindades D . En cuanto a la covariable inmigración vemos que es significativamente distinta de cero en un retardo de un año (en presencia del resto de covariables).

Este modelo explicaría la influencia directa de la mortalidad en regiones próximas a través de la componente auto-regresiva y la interacción simultánea indicándonos para cada localización qué vecinos en orden de proximidad y con qué retardo temporal influyen en mayor grado. Por una parte se refleja el posible intercambio y contacto entre poblaciones vecinas («próximas» geográficamente) y por otro la difusión espacial de la enfermedad a lo largo del tiempo.

La variable inmigración completaría la explicación anterior incorporando información sobre la llegada de población que no queda explicada por el contacto entre localizaciones vecinas («próximas» geográficamente) y la influencia se da en un retardo de un año.

No obstante, el estudio de los residuos del modelo nos lleva a observar que aún persisten errores sistemáticos en algunas localizaciones a lo largo del período (ver figura 3).

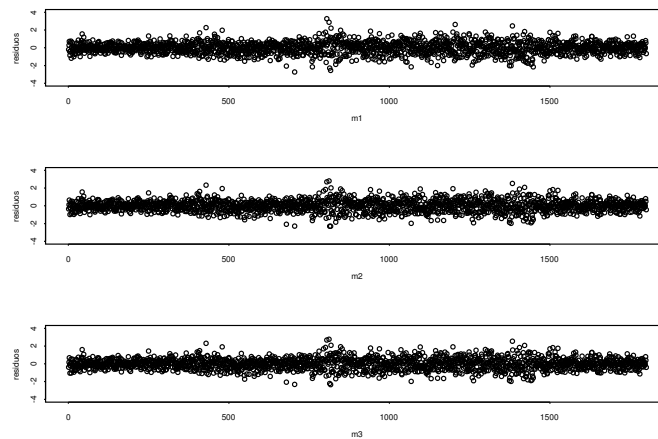


Figura 3. Residuos.

Para ver si el modelo recoge la dependencia espacial, consideramos los residuos como una serie ficticia univariante de longitud $(T - 5) * N$ (los 50 primeros términos corresponden al primer año, el segundo bloque de 50 elementos al segundo año y así sucesivamente ...), hemos visto que la función de autocorrelación de esta serie ficticia tiene un valor elevado en un retardo de 50, indicando que los residuos de una misma localidad en tiempos sucesivos siguen estando correlados. Este valor es de 0.2133 en el modelo 3.

A pesar de que este valor es menor que en los demás modelos (0.3341 para el modelo 1, y 0.2148 para el modelo 2), aún sigue siendo elevado (ver figura 4). Los residuos para los modelos 1, 2 y 3 (con dependencia espacial simultánea) aparecen en la figura 3.

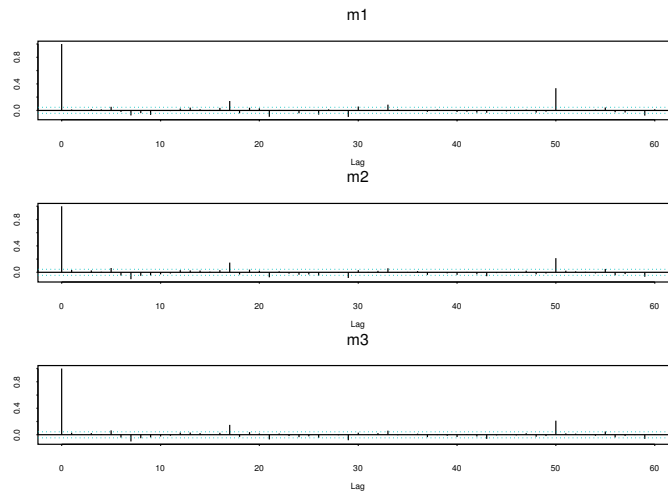


Figura 4. Función de autocorrelación para la serie ficticia de longitud $(T - s)N$.

5. CONCLUSIONES Y COMENTARIOS

En el presente trabajo hemos podido comprobar la adecuación de los auto-modelos dinámicos, en particular de la concatenación temporal de auto-modelos Gaussianos (CARX), para el análisis de datos geográficos de enfermedades infecciosas. Este tipo de modelo podría ser generalizado para abordar respuesta no Gaussiana mediante la consideración de otras distribuciones de la familia exponencial. Ello no entrañaría dificultad conceptual añadida en la etapa de modelización pero conllevaría una mayor complejidad en los algoritmos de maximización de la verosimilitud que, por cuanto conocemos hasta el momento, aún no están suficientemente perfeccionados.

En el estudio del caso particular de la mortalidad por meningitis en España, aceptamos el modelo 3 y con ello la hipótesis de una componente auto-regresiva espacio-temporal presente en el fenómeno estudiado así como una influencia de los movimientos migratorios en el comportamiento de la epidemia.

No obstante, el estudio de los residuos sugiere que el modelo debe ser mejorado. Como posible mejora podemos apuntar la naturaleza no lineal de la dependencia del propio pasado local, y/o el uso alternativo de otro tipo de auto-modelos no Gaussianos tales como el auto-Binomial o auto-Poisson (caso de partir de datos de mortalidad cruda). Este es un camino que aún requiere desarrollar nuevos procedimientos.

Otras posibles mejoras que pueden acometerse más fácilmente pasarían por la consideración de variantes de las matrices de dependencia espacial descritas en la expresión (9), que nosotros hemos especificado en función de la distancia geográfica pero que podrían basarse en otros criterios alternativos.

Finalmente la hipótesis de tendencia local temporal puede afectar también los resultados del análisis espacio-temporal.

6. AGRADECIMIENTOS

Los autores quieren agradecer los comentarios y sugerencias de los evaluadores, los cuales han contribuido a la mejora del artículo. Este trabajo ha sido parcialmente financiado por los proyectos PB96-0776 de la Dirección General de Investigación Científica y Técnica (DGICYT) y GV99-621-01 de la Generalitat Valenciana.

REFERENCIAS

- Becker, R., Chambers, J. y Wilks, A. (1988). *The New S Language. A Programming Environment for Data Analysis and Graphics*. Wadsworth and Brooks/Cole, California.
- Besag, J. (1974). «Spatial interaction and the statistical analysis of lattice systems». *Journal of the Royal Statistical Society B*, 36, 192-225.
- Cliff, A. D. y Hagget, P. (1993). «Statistical modelling of measles and influenza outbreaks». *Statistical Methods in Medical Research*, 2, 43-73.
- Cliff, A. D., Hagget, P. y Stroup, D. F. (1992). «The geographic structure of measles epidemics in the northeastern united states». *American Journal of epidemiology*, 136, 592-602.
- Cliff, A. D., Ord, J. K., Basset, K. y Davies, R. B. (1975). *Elements of spatial structure: a quantitative approach*. Cambridge University Press, London.

- Cook, D. y Pocock, S. (1993). «Multiple regression in geographical mortality studies, with allowance for spatial correlation errors». *Biometrics*, 39, 361-371.
- Cressie, N. (1991). *Statistics for Spatial Data*. Wiley, New York.
- Darroch, J., Lauritzen, S. y Speed, T. (1980). «Markov fields and log-linear interaction models for contingency tables». *The Annals of Statistics*, 8, 522-539.
- Ferrándiz, J., López, A., Morales, M. y Tejerizo, M. (1995). «Spatial interaction between neighbouring counties: cancer data in Valencia (Spain)». *Biometrics*, 51, 665-678.
- Ferrándiz, J., López, A. y Sanmartín, P. (1999). «Spatial regression models in epidemiological studies». En *Disease Mapping and Risk Assessment for Public Health* (Lawson, A., Boehning, D., Lesaffre, E., Biggeri, A., Viel, J. F. y Bertollini, R., eds.), 203-215, Wiley, Chichester.
- Fisher, M. y Getis, A. (1997). *Recent developments in Spatial Analysis*. Springer, Berlin.
- Geman, D. y Geman, S. (1984). «Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images». *IEEE transactions of pattern Analysis and Machine Intelligence*, PAMI-6, 721-741.
- Gibbs, W. (1902). *Elementary principles of statistical mechanics*. Yale University Press, Connecticut.
- Knorr-Held, L. y Besag, J. (1998). «Modelling risk from disease in time and space». *Statistics in Medicine*, 17, 2045-2060.
- Logan, J. y Cliff, O. (1997). «A structural time series approach to forecasting the space-time incidence of infectious diseases: Post-war measles elimination programmes in the united states and iceland». En *Recent developments in Spatial Analysis* (Fisher, M. y Getis, A. eds.), 101-127, Springer, Berlin.
- López, A. (1997). *Modelos Lineales Generalizados Espaciales*. Tesis doctoral, Departament d'Estadística i I.O. (Universitat de València).
- Mardia, K. V. y Marshall, R. J. (1984). «Maximum likelihood estimation of models for residual covariance in spatial regression». *Biometrika*, 71, 135-146.
- Mardia, K. V. y Watkins, A. J. (1989). «On multimodality of the likelihood in the spatial linear model». *Biometrika*, 76, 289-295.
- Martin, R. y Oeppen, J. (1975). «The identification of regional forecasting models using space-time correlation functions». *Transactions of the Institute of British Geographers*, 66, 95-118.
- Martinez, F. (1989). «Reproducción social y enfermedad el caso de la infección meningocócica en España». *Actas del sexto Congreso Mundial de Medicina Social*.
- McCullagh, P. y Nelder, J. (1989). *Generalized linear models*. Chapman and Hall, London.

- Pfeifer, P. y Deutsch, S. (1980). «A three-stage iterative procedure for space-time modelling». *Technometrics*, 22, 35-47.
- Press, W., Flannery, B., Teukolsky, S. y Vetterling, W. (1986). *Numerical Recipes, the Art of Scientific Computing*. Cambridge University Press, Cambridge.
- Richardson, S., Guihenneuc, C. y Lasserre, V. (1992). «Spatial linear models with auto-correlated error structure». *The Statistician*, 41, 539-557.
- Sanmartín, P. (1997). *Agregación temporal en modelos de grafos cadena*. Tesis doctoral, Departament d'Estadística i I.O. (Universitat de València).
- Stoffer, D. (1986). «Estimation and identification of space-time ARMAX models in the presence of missing data». *Journal of the American Statistical Association*, 81, 762-772.
- Venables, W.N. y Ripley, B. D. (1999). *Modern Applied Statistics with S-plus*. Springer-Verlag, New York.
- Waller, L., Carlin, B. P., Xia, H. y Gelfand, A. E. (1997). «Hierarchical spatio-temporal mapping of disease rates». *Journal of the American Statistical Association*, 92 (438), 607-617.
- Warnes, J. J. y Ripley, B. D. (1987). «Problems with likelihood estimation of covariance functions of spatial Gaussian processes». *Biometrika*, 74, 640-642.

ENGLISH SUMMARY

TEMPORAL CONCATENATION OF SPATIAL MODELS AND ITS APPLICATION TO THE STUDY OF MENINGITIS IN SPAIN

JUAN FERRÁNDIZ FERRAGUD*
FERRAN MARTÍNEZ I NAVARRO**
PILAR SANMARTÍN FITA*

Disease mapping of infectious diseases leads in a natural way to its dynamical consideration. This work undertakes the temporal concatenation of spatial auto-regressive models to perform the analysis of meningitis mortality in Spain for the period 1950-1990 from data aggregated at province scale. Likelihood-based methods are used to achieve parameter estimation and model selection.

Keywords: Spatial-temporal analysis, spatial auto-regressive models, infectious diseases

AMS Classification (MSC 2000): 62M10, 62M30, 62P10

* Departament d'Estadística i I.O., Universitat de València, València.

** Instituto de Salud Carlos III, Madrid.

* Departamento de Matemática Aplicada y Estadística, Universidad Politécnica de Cartagena, Paseo Alfonso XIII s/n, 30203 Cartagena.

– Received June 1999.

– Accepted January 2001.

1. INTRODUCTION

Geographical studies of epidemiological data over time require space-time regression models in order to capture the influence of covariates and spatial interactions along the studied period. Hierarchical models have been successfully applied in disease mapping thanks to MCMC methods. They are being extended to cope with spatio-temporal analysis of mortality/morbidity geographical data [see e. g. Waller et al. (1997); Knorr-Held and Besag, (1998)].

Hierarchical (or multilevel) models consider spatial interactions in a second stage of the model thus providing a natural way to incorporate the effect of hidden environmental factors.

Nevertheless, when considering infectious diseases, a direct influence of mortality/morbidity values between neighbouring regions is expected. Spatial auto-regressive models, such as Besag's auto-models [Besag (1974)], seem a more natural and intuitive approach.

In this work we undertake the analysis of meningitis mortality data in Spain by linking spatial auto-models in time. Population immigration effect has been taken into account as a covariate.

Section 2 introduces the model. Fitting methods are discussed in section 3. Section 4 is devoted to the particular analysis of annual meningitis mortality rates in Spain for the period 1950 – 1990 and the 50 peninsular provinces. Finally in section 5 we give some results and conclusions.

2. MODEL

Let us consider a multivariate time series $\{\mathbf{y}(t)\}_{t=1}^T$ where, for any time t , $\mathbf{y}(t) \in \mathbb{R}^N$ represents the data collected at sites $i = 1, \dots, N$. Given the spatial structure of the aimed data we approach the problem as a temporal concatenation of spatial Gibbs distributions. We consider the class of Besag's auto-models [Besag, (1974)] in general and the conditionally-specified auto-Gaussian model with covariates (CARX) [see Cressie (1991); Richardson et al (1992)] in particular. Their extension to the dynamic framework has been studied in [Sanmartín (1997)] according to the following formulation:

$$(1) \quad p(y_i(t) | \mathbf{y}(t)_{-i}, \mathbf{y}(t)^*, \mathbf{x}_i(t), \mathbf{x}_i(t)^*) \propto \exp\left\{a_i(t) + \sum_{j \in \delta_i} \beta_{ij}(t) y_j(t) y_i(t) + h_i(y_i(t))\right\}.$$

where $\mathbf{y}(t)_{-i}$ stands for the vector $\mathbf{y}(t)$ excluding its i -th component, and $\mathbf{y}(t)^*$, $\mathbf{x}(t)^*$,

stand for the past of $\mathbf{y}(t)$ and $\mathbf{x}(t)$ respectively. These variables enter the model through the term

$$a_i(t) = a_i(\mathbf{y}(t)^*, \mathbf{x}_i(t), \mathbf{x}_i(t)^*)$$

in the same way as covariates do for the static case in Ferrándiz et al. (1995) and López, (1997). In the auto-Gaussian model case, (1) leads to

$$(2) \quad \mathbf{y}(t) \mid X(t), \mathbf{y}(t)^*, X(t)^* \sim N(\boldsymbol{\mu}(t), (I - B(t))^{-1} \text{Diag}(\boldsymbol{\sigma}_i^2(t)))$$

where the vector of means $\boldsymbol{\mu}(t)$ depends on $\mathbf{x}(t)$, $\mathbf{y}^*(t)$, $\mathbf{x}^*(t)$ according to:

$$(3) \quad \boldsymbol{\mu}(t) = \sum_{i=1}^p A_{0i} \mathbf{y}(t-i) + \sum_{j=0}^{q_r} A_{rj} \mathbf{x}(t-j)$$

Matrices A_{0i} and A_{rj} reflect spatial dependence and are usually expressed as linear combinations of fixed matrices [Cressie (1991); Stoffer (1981)]. We also consider this simplification throughout the paper. Moreover, $B(t) \equiv B$ and $\boldsymbol{\sigma}_i^2(t) \equiv \boldsymbol{\sigma}^2$ are considered independent with respect the time and time/location respectively.

3. MODEL FITTING

Let $s = \max\{p, q_1, \dots, q_r\}$ be the maximum temporal lag in (2). Conditioning to the s initial times, the log-likelihood function becomes

$$(4) \quad \begin{aligned} \ell(\boldsymbol{\mu}, B, \boldsymbol{\sigma}^2) &= \log\left(\prod_{t=s+1}^T p(\mathbf{y}(t) \mid X(t), \mathbf{y}(t)^*, X(t)^*)\right) \\ &= -\frac{1}{2\boldsymbol{\sigma}^2} Q(B, \boldsymbol{\mu}) + \frac{T-s}{2} \log |I - B| - \\ &\quad - \frac{(T-s)N}{2} \log \boldsymbol{\sigma}^2 + \text{const.} \end{aligned}$$

where

$$(5) \quad Q(B, \boldsymbol{\mu}) = \sum_{t=s+1}^T (\mathbf{y}(t) - \boldsymbol{\mu}(t))' (I - B) (\mathbf{y}(t) - \boldsymbol{\mu}(t))$$

Maximum likelihood estimators can be obtained by maximizing the *profile* log-likelihood of B :

$$(6) \quad \hat{B} = \underset{B}{\text{argmax}} \ell(B, \hat{\boldsymbol{\mu}}(B), \hat{\boldsymbol{\sigma}}^2(\hat{\boldsymbol{\mu}}, B))$$

where

$$\hat{\boldsymbol{\mu}}(B) = \underset{\boldsymbol{\mu}}{\operatorname{argmin}} Q(B, \boldsymbol{\mu})$$

$$\hat{\sigma}^2(B, \hat{\boldsymbol{\mu}}(B)) = \frac{1}{(T-s)N} Q(B, \hat{\boldsymbol{\mu}}(B))$$

are the usual solution to the corresponding generalized least-squares problem for fixed B , as it is done in Cressie (1991) and Richardson et al. (1992) for the static CARX model.

Maximization of the profile likelihood can be performed by means of numerical robust procedures such as those based on the *downhill simplex method* of Nelder and Mead [see Press et al. (1986)].

4. MENINGITIS MORTALITY RATES IN SPAIN (1950-1990)

Here we analyze meningitis mortality annual rates in the spanish provinces (excluding Ceuta and Melilla), for the period 1950 to 1990. We have considered a temporal lag $p = 5$. The spatial matrices A_{rj} have been built up as linear combinations of constant matrices D_k corresponding to three degrees of geographic proximity. The immigration population effect has been incorporated as a covariate with temporal lag $q_1 = 1$. The following models have been compared:

Model 1 Different Temporal polynomial cubic trend for each site.

Model 2 Auto-regressive spatial-temporal model with local temporal trend (temporal order 5, spatial order 3).

Model 3 Auto-regressive spatial-temporal model with local temporal trend (temporal order 5, spatial order 3) with immigration covariate (temporal lag 1, only local influence).

We have compared these models by using the likelihood ratio test with the corresponding asymptotic χ^2 distributions.

5. COMMENTS AND CONCLUSIONS

The proposed dynamic extension of auto-Gaussian models has proved to be appropriate in order to perform spatial-temporal analysis of infectious diseases geographical data. It seems a sensible alternative in which spatial contagion interactions can be incorporated in a natural way. Its statistical analysis can be performed by likelihood-based methods.

In the particular analysis of meningitis mortality rates in Spain, **Model 3** has been proved to be the best and, consequently, beyond the implied temporal auto-regressive terms, a simultaneous spatial interaction has to be assumed. Furthermore, the influence of population immigration movements is significant in the evolution of the epidemic.

The residual study shows that this model still needs some improvement. Non-linear local temporal auto-regression and the resource to other discrete auto-models such Binomial or Poisson could overcome these difficulties.

6. ACKNOWLEDGEMENTS

The authors want to thank the referees for comments and suggestions. This work has been partially supported by the Dirección General de Investigación Científica y Técnica (DGICYT) and Generalitat Valenciana (projects PB96-0776 and GV99-621-01).