

# UNA VISIÓN INTEGRAL DEL ESPACIO Y EL TIEMPO

A. López-Quílez

Departament d'Estadística i Investigació Operativa, Universitat de València

## Introducción

Este trabajo presenta en clave de síntesis el esfuerzo desarrollado en el Grupo de Estadística Espacial y Temporal en Epidemiología y Medio Ambiente (GEE<sub>t</sub>•E<sub>ma</sub>, <http://matheron.estadi.uv.es/geeitema/>) durante más de una década. Los estudios parciales realizados en diferentes temas, empleando distintos modelos y técnicas, pueden verse bajo una perspectiva amplia del análisis espacial y temporal proporcionando una visión de conjunto.

El empleo de técnicas estadísticas en Epidemiología es habitual desde sus inicios. Recientemente se ha acuñado el término Epidemiología Espacial referido a diferentes tópicos sobre el estudio de la dispersión espacial de enfermedades, incluyendo la cartografía de enfermedades, la detección de agrupaciones de casos, el análisis ecológico, etc. El interés creciente en esta materia ha motivado la aparición de números especiales en revistas científicas prestigiosas y la publicación de libros monográficos (Lawson et al., 1999; Elliot et al., 2000; Lawson, 2001; Briggs et al., 2002; Haining, 2003).

Conforme avanzan los años, parecen aumentar las posibilidades para analizar estadísticamente problemas epidemiológicos de gran complejidad, con datos multivariantes de naturaleza espacial y temporal. Los *métodos de Monte Carlo por cadenas de Markov* (MCMC) proporcionan una herramienta para analizar estas situaciones que ha propiciado el rápido desarrollo de la Estadística Espacial desde la última década del siglo XX. A pesar de ello, todavía quedan muchas cuestiones pendientes, como la gran complejidad de los modelos o la fiabilidad de las técnicas empleadas.

El tiempo es una componente importante en cualquier estudio científico dinámico. Así, la modelización del desarrollo y evolución de epidemias requiere de modelos espacio-temporales, pero también otros estudios deben incorporar la tendencia temporal como expresión de su carácter dinámico. El análisis de series temporales o la introducción de factores diferenciadores del comportamiento en el tiempo son elementos habituales en el estudio epidemiológico.

La variabilidad espacial está presente también en cualquier investigación epidemiológica. La decisión crucial a tomar es entre modelizar la variabilidad espacial o eliminarla. Desafortunadamente, en muchos estudios no es posible realizar un riguroso diseño del experimento. La situación es compleja cuando estudiamos personas y la influencia de su entorno. En estudios epidemiológicos, se pueden seleccionar controles equilibrados con los casos, pero los casos no son asignados aleatoriamente a las dis-

tintas zonas. Además, una investigación de los efectos de un factor de riesgo sobre la salud no admite la replicación.

Así pues, en vez de un cuidadoso diseño de experimentos, nos enfrentamos a problemas planteados sobre estudios observacionales. A menudo una única observación constituye toda la información disponible en cada zona geográfica por unidad de tiempo.

El principal objetivo de la comunicación es presentar una idea de conjunto de los modelos y las técnicas espaciales y temporales desde una perspectiva jerárquica. En la siguiente sección se identifican los tres tipos de datos espaciales habituales, mencionando ejemplos de estudios epidemiológicos realizados. El resto de la comunicación está dedicada a la integración de los análisis espaciales y temporales en un marco amplio de modelos jerárquicos bayesianos.

## Datos espaciales

Los problemas epidemiológicos con frecuencia vienen relacionados con observaciones espaciales de distinta índole. Los datos son continuos o discretos, están agregados espacialmente o son observaciones individuales en puntos del espacio, sus localizaciones se encuentran dispuestas de forma regular o irregular, e incluso, estas localizaciones provienen de una región espacial continua o de un conjunto discreto.

Los datos espaciales se pueden clasificar en tres grupos fundamentales según el contexto de observación del que provienen: observaciones de un fenómeno continuo en el espacio, datos en una red fija de localizaciones y sucesos que ocurren en el espacio proporcionando un conjunto aleatorio de puntos llamado patrón puntual. Estos tipos de datos diferenciados dan origen a formas distintas de modelización y, por tanto, de análisis estadístico (Cressie, 1993).

### Geoestadística.

La *Geoestadística* es un término acuñado para denominar a las técnicas estadísticas aplicadas al análisis geográfico. Hay una gran variedad de problemas que pueden resolverse utilizando métodos geoestadísticos. La característica común a todos ellos es que los datos pueden verse como una realización, habitualmente parcial, de un proceso estocástico sobre una región espacial continua.

La clave fundamental en la modelización de la relación espacial en el proceso es el variograma que es objeto de modelización y estimación para describir adecuadamente el fenómeno observado.

El objetivo principal en la aplicación de la geoestadística es habitualmente la predicción en un punto o en un conjunto de puntos de la región observada. La técnica de predicción espacial más empleada es el *kriging* (Goovaerts, 1997).

Abellán et al. (2002b) aplican el kriging a la modelización de la tasa de incidencia de gripe haciendo uso de la red centinela de la Comunidad Valenciana.

#### Datos en áreas pequeñas.

Estudio de situaciones en que las observaciones provienen de un conjunto fijo de localizaciones. La predicción en otros puntos del espacio no tiene sentido cuando el fenómeno observado únicamente ocurre en localizaciones fijas o cuando es observado agregadamente en pequeñas áreas.

Los modelos estadísticos para este tipo de datos tienen que expresar el hecho de que las observaciones próximas tienden a ser parecidas. Por tanto, deben incorporar la relación existente entre las observaciones o los riesgos de localizaciones vecinas. La especificación de estas relaciones a partir de las distribuciones condicionales origina los campos aleatorios markovianos (Besag, 1974).

Ferrándiz et al. (1995) estudian el efecto de la concentración de nitratos en el agua potable sobre la mortalidad por diferentes tipos de cáncer empleando un modelo autoregresivo sobre las observaciones municipales.

#### Patrones puntuales.

Un patrón puntual es una colección de ubicaciones de fenómenos que ocurren aleatoriamente en diferentes puntos de una región. Cada uno de estos puntos puede representar un objeto de pequeño tamaño en comparación con la región observada, la aparición de un determinado fenómeno o la ocurrencia de un suceso asociado a ese lugar geográfico (Diggle, 2003).

Los procesos puntuales permiten establecer diferentes mecanismos teóricos que modelizan la formación de estos fenómenos. Las características de esos procesos puntuales proporcionan información sobre la estructura de formación de los sucesos.

Abellán et al. (2002a) realizan el análisis de casos y controles en un brote de neumonía por Legionella en la ciudad de Alcoi mediante técnicas de procesos puntuales.

## Modelización integral

La modelización estadística de los problemas epidemiológicos conlleva la adecuada incorporación de estos fenómenos y de las relaciones de dependencia espacial y temporal que puedan existir entre las observaciones. Diferentes aproximaciones se han empleado para los tres tipos de datos espaciales, todas ellas inspiradas en aspectos del análisis temporal (Ripley, 1981).

Pero la incorporación al modelo de componentes espaciales y temporales tiene elementos comunes que conviene identificar y relacionar entre sí. Esto queda más

patente con la evolución de estos tres ámbitos hacia los modelos jerárquicos bayesianos (Banerjee et al., 2004).

Los modelos jerárquicos añaden cierta complejidad a los modelos tradicionales permitiendo la construcción de estructuras flexibles mediante el encadenamiento condicional de modelos simples. Están suponiendo la principal herramienta de análisis en problemas en que varios fenómenos están interconectados o que tienen variables de muy diversa índole. Su auge viene dado por su capacidad de adaptarse a situaciones complejas y por el desarrollo de técnicas inferenciales asequibles mediante simulación (Møller, 2003).

El uso de modelos jerárquicos bayesianos se ha generalizado en los tres tipos de datos espaciales. Primero se extendió el modelo de Besag et al. (1991) para el problema de suavización de mapas de riesgo de enfermedad o cartografía de enfermedades en áreas pequeñas. Recientemente ha irrumpido la generalización de la geoestadística con el trabajo de Diggle et al. (1998). Y de forma todavía embrional los procesos de Cox log-gaussianos introducidos por Møller et al. (1998) para analizar datos puntuales.

Estos tres modelos jerárquicos comparten elementos que antes parecían exclusivos de cada parcela. Uno de ellos es la estructura de respuesta generalizada, habitualmente de tipo Poisson, en la primera capa de las observaciones, modelizando de forma lineal el logaritmo de su media. Otro elemento común es la inclusión en la segunda capa de las componentes espaciales y temporales autoregresivas, con relaciones marcadas por el variograma en términos de la distancia entre las observaciones. También ayuda la concepción de descomposición de la variabilidad a gran escala y a pequeña escala, a imagen del análisis de series temporales, que distingue la tendencia global de las variaciones locales. Todos estos modelos permiten introducir la explicación aportada por las covariables y la heterogeneidad captada mediante efectos aleatorios.

Los esfuerzos realizados en distintos problemas pueden trasladarse a otros contextos mediante la identificación del papel de estos elementos comunes. Además el paso de una técnica a otra en un mismo problema es posible si el tipo de datos recogidos varía por agregación o ausencia de observaciones. Todo ello conduce a una modelización única de los problemas epidemiológicos espaciales y temporales que surge de una visión integral del espacio y el tiempo.

## Agradecimientos

Este trabajo ha sido parcialmente financiado por la Dirección General de Salud Pública de la Generalitat Valenciana mediante un acuerdo de colaboración con la Universitat de València.

Esta investigación ha sido desarrollada por el **GE<sub>7</sub>t•E<sub>7</sub>ma** compuesto por Carlos Abellán, Juanjo Abellán, Carmen Armero, María Jesús Bayarri, Paloma Botella, David Conesa, Virgilio Gómez-Rubio, Rut López, Miguel Ángel Martínez-Beneito, Jordi

Pérez-Panadés, Pilar Sanmartín y, por supuesto, Juan Ferrándiz, quien nos condujo por estos derroteros, nos enseñó e impulsó, transmitiendo su visión y ánimo.

## Referencias

- Abellán, J. J., Martínez-Beneito, M., Zurriaga, O., Jorques, G., Ferrándiz, J. y López-Quílez, A. (2002a). Procesos puntuales como herramienta para el análisis de posibles fuentes de contaminación. *Gaceta Sanitaria* 16:445–449.
- Abellán, J. J., Zurriaga, O., Martínez-Beneito, M. A., Peñalver, J. y Molins, T. (2002b). Aplicación de la metodología geoestadística a la vigilancia de la gripe en una red centinela. *Gaceta Sanitaria* 16:324–333.
- Banerjee, S., Carlin, B. P. y Gelfand, A. E. (eds.) (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall, Boca Raton.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, series B* 36:192–236.
- Besag, J., York, J. C. y Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics* 43:1–59.
- Briggs, D. J., Forer, P., Järup, L. y Stern, R. (eds.) (2002). *GIS for Emergency Preparedness and Health Risk Reduction*. Kluwer Academic Publishers.
- Cressie, N. A. (1993). *Statistics for Spatial Data*. New York: John Wiley and Sons, segunda edición.
- Diggle, P. J. (2003). *Statistical Analysis of Spatial Point Patterns*. London: Arnold., Segunda edición.
- Diggle, P. J., Tawn, J. A. y Moyeed, R. A. (1998). Model-based geostatistics (con discusión). *Applied Statistics* 47:299–350.
- Elliot, P., Wakefield, J. C., Best, N. G. y Briggs, D. J. (2000). *Spatial Epidemiology: Methods and Applications*. Oxford University Press.
- Ferrándiz, J., López, A., Llopis, A., Morales, M. y Tejerizo, M. (1995). Spatial interaction between neighbouring counties: Cancer mortality data in Valencia (Spain). *Biometrics* 51:665–678.
- Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York.

- Haining, R. (2003). *Spatial Data Analysis. Theory and Practice*. Cambridge University Press.
- Lawson, A. B. (2001). *Statistical Methods in Spatial Epidemiology*. Wiley.
- Lawson, A. B., Biggeri, A., Böhning, D., Lesaffre, E. y Viel, J. F. (eds.) (1999). *Disease Mapping and Risk Assessment for Public Health*. John Wiley and Sons Ltd.
- Møller, J. (ed.) (2003). *Spatial Statistics and Computational Methods*. Springer Verlag, New York.
- Møller, J., Syversveen, A. R. y Waagepetersen, R. P. (1998). Log gaussian cox processes. *Scandinavian Journal of Statistics* 25:451–482.
- Ripley, B. D. (1981). *Spatial Statistics*. John Wiley and Sons, New York.