

8. Modelización de procesos con evolución en espacio-tiempo

Dos son las líneas de trabajo que actualmente se desarrollan en este contexto.

Línea de investigación 1: *Nuevas familias de covarianzas espacio-temporales no separables, no estacionarias y anisotrópicas: Marco teórico y aplicaciones.*

En los últimos años la demanda de modelos que describen la evolución de procesos medioambientales en espacio y tiempo ha sido muy considerable. En particular, hay una clara necesidad de modelos que capturen el comportamiento espacial y temporal de forma simultánea. En este contexto, es la geoestadística espacio-temporal la que nos ofrece los instrumentos adecuados de análisis estadístico. Como es sabido, los procedimientos de Kriging basados en interpolación dependen fuertemente de la elección de la estructura de autocovarianza asociada al campo aleatorio espacio-temporal. Por tanto, el objetivo primordial de la geoestadística es obtener estructuras permisibles de covarianzas espacio-temporales que consideren las dependencias e interacciones entre el espacio y el tiempo.

Por otra parte, la estacionariedad constituye una de las hipótesis más comunes para el desarrollo de modelos y herramientas estadísticas en el análisis de datos, especialmente en el caso espacio-temporal, donde la dimensionalidad del problema obliga en muchos casos a la consideración de hipótesis de este tipo (junto con la isotropía espacial). Sin embargo, dicha hipótesis no es realista en muchas ocasiones. Si a esto añadimos que hay una necesidad creciente de modelar datos espacio-temporales bajo presencia de anisotropía y en general de no estacionariedad, pero al mismo tiempo muchas de las propuestas de modelos de covarianza espacio-temporal no separables no pueden ser utilizadas para tratar el problema de la anisotropía, nos encontramos con un ‘gap’ que hay que rellenar. En esta línea de trabajo nos proponemos rellenar esta gap desde el ámbito de la Teoría de Procesos Estocásticos.

Las propuestas iniciales de modelos de covarianza espacio-temporal se centraban en estructuras separables y estacionarias. Posteriormente, se trataron las estructuras separables en combinación con modelos anisotrópicos. En particular: (a) Dimitrakopoulos & Lou (1994) propusieron algunos modelos geométricamente anisotrópicos para datos espacio-temporales; (b) Guttorp et al. (1992) propusieron algunas estructuras de covarianza separables a través del producto tensorial de una covarianza espacial con otra temporal; (c) Rohuani & Hall (1989) propusieron el variograma suma; (d) Otros acercamientos se basaron en el hecho de eliminar la variabilidad temporal a través de una tendencia, para conseguir un proceso puramente espacial (Sampson & Guttorp, 1992). Las estructuras no separables, aunque todavía estacionarias e isotrópicas en espacio y tiempo, llegaron de la mano de Jones & Zhang (1997), Cressie & Huang (1999), Christakos (2000), De Cesare et al. (2001), Gneiting (2002a), Ma (2002, 2005), Stein (1999, 2005) y Fernández-Casal et al. (2003). En particular: (a) Cressie & Huang (1999) propusieron un acercamiento espectral, más adelante generalizado por Gneiting (2002a) mediante el uso de funciones completamente monótonas y

funciones cuya primera derivada es completamente monótona; (b) Stein (2005) pone el énfasis en el desarrollo espectral de la función de covarianza y en el hecho de que una densidad espectral espacio-temporal debe ser suficientemente suave lejos del origen; (c) Fernández-Casal et al. (2003) extienden el método de Shapiro-Botha para obtener variogramas espacio-temporales suficientemente flexibles. Fue al principio del 2000 cuando algunos autores se atreven con estructuras no estacionarias. Cabe destacar Christakos (2002), Fuentes & Smith (2001), Fuentes (2002) y Kolovos et al. (2004). Sin embargo, muchas de las estructuras propuestas tienen ciertos inconvenientes (poco flexibles, difícil interpretación, complicadísima computación en el análisis práctico de los modelos) que impiden, y de hecho han impedido hasta hoy en día, ser consideradas estructuras adecuadas para el análisis de dependencias espacio-temporales.

Algunos de los objetivos de esta línea de trabajo son los siguientes:

1) Objetivo: Modelos anisotrópicos y no estacionarios 1.1) Construcción de modelos no estacionarios para datos espacio-temporales a través de: (a) por una representación adecuada del proceso subyacente; (b) por construcción directa de la función de covarianza; (c) por construcción de las correspondientes densidades espectrales y transformadas de Fourier. 1.2) Proposición de nuevas clases de funciones de covarianzas estacionarias y no separables que puedan ser utilizadas en problemas de anisotropía zonal y geométrica. 1.3) Búsqueda de modelos para datos espacio-temporales no estacionarios. Para ello nos adentraremos en el terreno del “spatial adapting” “spectral tempering” sobre densidades espectrales para obtener funcionales arquimedeanos generados por funciones completamente monótonas.

2) Objetivo: Funciones de correlación sobre soporte compacto. La importancia de las funciones de correlación sobre soporte compacto es que presentan una serie de ventajas desde el punto de vista computacional, permitiendo una más rápida computación de matrices inversas e implementación del sistema de Kriging. En particular, haremos uso de la función de Golubov para construir funciones de covarianza no estacionaria y que eventualmente pueden tomar valores negativos.

3) Objetivo: covarianzas basadas en medias cuasi-aritméticas Establecimiento de conexiones entre las medias cuasi-aritméticas y las funciones definidas positivas. En particular: 3.1) Podemos utilizar las medias cuasi-aritméticas para establecer criterios de permisibilidad para funciones definidas en espacios euclídeos $\mathbb{R}^d \times \mathbb{R}$. 3.2) Podemos utilizar las medias cuasi-aritméticas para un mejor conocimiento de algunas propiedades de los campos aleatorios (débilmente) estacionarios. Como en el caso precedente, la respuesta es afirmativa pero es preciso analizarla en profundidad. 3.3) Completar el cuadro de las propiedades de funciones de covarianza. 3.4) Hacia un mejor cuadro de las propiedades campos aleatorios estacionarios. 3.5) Cuasi-aritmeticidad y estacionariedad local.

4) Objetivo: estimación empírica de estos tipos de estructuras y análisis de datos reales.

Proponemos el método de la “Composite Likelihood” como alternativa natural y adecuada a la verosimilitud y métodos Least Squares. Esto supone un avance importante pues los métodos de Composite sólo han sido desarrollados hasta el momento para datos espaciales pero no para problemas espacio-

temporales.

Línea de investigación 2: *Modelización de interdependencias espacio-tiempo en procesos puntuales.*

Esta línea de trabajo analiza las posibles interdependencias entre espacio y tiempo en procesos puntuales mediante el uso de procesos estocásticos espacio-temporales. De forma particular se persiguen los siguientes objetivos:

- (a) Dinámica espacio-temporal mediante funciones de intensidad.
- (b) Procesos espaciales birth-death a través de funciones de intensidad.
- (c) Clustering espacial sobre configuraciones estructuradas.
- (d) Movimientos de partículas aleatorios y Gibbsianos.

8.0.1. Didáctica de las Matemáticas

El primer objetivo actual del grupo de investigación consiste en aprovechar las cualidades generales del tratamiento de la información, utilizando instrumentos comunes de análisis -el análisis tipológico y el análisis implicativo- en campos diversos. Fundamentalmente en el ámbito de la formación por el importante trabajo cognitivo que potencia (lógico, clasificadorio y deductivo) asociado al análisis de datos y en otros ámbitos, como puede ser la investigación en Didáctica de las Matemáticas, puede como instrumento de ayuda en la toma de decisiones basadas en la información. La enseñanza secundaria será el ámbito para el diseño y la experimentación de situaciones didácticas basadas en el análisis de tablas de datos y en el tratamiento gráfico de la información.

El segundo objetivo consiste en seguir utilizando y difundiendo estas técnicas innovadoras de tratamiento de la información tanto en sectores no especializados, tales como alumnos de la ESO, profesores de enseñanza obligatoria y universitaria, empresarios, etc., como en los sectores especializados en estas nuevas técnicas de tratamiento estadístico de datos multivariantes y extracción de conocimientos, como son el Análisis Clasificadorio (Lerman; 1970, 1995) basado en la obtención de clases de datos agrupados por similaridad, el Análisis Implicativo y Cohesivo (Gras y Coll; 1992, 1996, 2000) y el Tratamiento Gráfico de dichos resultados, mediante árboles y grafos -con la utilización del programa informático C.H.I.C (Classification Hiérarchique Implicative et Cohésitive) (Gras, Couturier, versión del 2005)-, y la semiología gráfica (Bertin, 1977). - Por último se pretende realizar nuevas aportaciones metodológicas respecto a la utilización de estos tipos de análisis de datos también en sectores más especializados en análisis estadísticos, tanto respecto a aspectos concretos de estos tipos de análisis en sí mismos, como respecto a su aplicación en investigación en Didáctica de las Matemáticas. En resumen, es esta potencialidad instrumental y didáctica que nos ofrece el tratamiento de datos multivariantes, el núcleo de la investigación actual del grupo que dará lugar a líneas de investigación diferentes pero interrelacionadas entre sí, como se plasmó en el proyecto "Innovaciones tecnológicas en el tratamiento de la información y aportaciones metodológicas en el ámbito de la investigación" (Bancaixa-UJI. 2002/05)

Línea de investigación 1: *Lógica, tratamiento de datos y estadística en la enseñanza obligatoria*

En esta línea se pretende utilizar el análisis tipológico (clasificadorio) como instrumento didáctico, para que los alumnos de la ESO, puedan realizar el tratamiento de los datos y de la lógica simultáneamente, a partir de la introducción del análisis estadístico (correspondiente al nivel de esta etapa educativa). En este sentido se continúan los trabajos realizados por P. Orús (desde su tesis, en 1992) sobre la lógica y análisis de datos en la enseñanza primaria en Francia. Introducir en la ESO el tratamiento de datos no es una novedad. Los contenidos curriculares contemplan este bloque temático, si bien en la realidad, generalmente se reduce a la enseñanza de conceptos estadísticos básicos de representación de datos, si bien el título del bloque temático sigue siendo tratamiento de datos y sus objetivos son bastante más ambiciosos que lo enseñado en clase. Prestigiosos investigadores en Educación estadística señalan la importancia y las posibilidades didácticas de este tema que compartimos, existiendo algunas experiencias en este tema (Batanero 1991, 1999), y las que podemos encontrar, recogidas en el número monográfico “Tratamiento de la información” de la Revista de Didáctica de las Matemáticas UNO. No obstante estos mismos artículos muestran la insuficiencia y las limitaciones de los trabajos actuales, y las posibilidades que aporta nuestro proyecto: el análisis tipológico (clasificadorio) como instrumento didáctico, para que los alumnos de la ESO puedan realizar el tratamiento de los datos y de la lógica simultáneamente, a partir de la introducción del análisis estadístico (correspondiente al nivel de esta etapa educativa). En su trabajo de investigación de tercer ciclo, Pitarch (2002), “Estudio sobre la viabilidad e interés didáctico del tratamiento de la información en la ESO”, ha realizado una primera aproximación experimental al tema y muestra el interés didáctico que puede tener el trabajo del tratamiento de datos en la enseñanza secundaria, trabajo que se completa en (Pitarch, Orús, 2003 y 2004). Este trabajo nos hace conjeturar que la ingeniería didáctica que se busca para alumnos de la ESO es interesante didácticamente, y que basándonos en el tipo de trabajo sobre las tablas de datos propuesto por Pitarch y Orús (2004) esa ingeniería también es viable. En la actualidad, se están elaborando las tesis respectivas, a partir del diseño y experimentación de las situaciones didácticas que implementan este trabajo conjunto -lógico, clasificadorio, estadístico y deductivo-, sobre soportes distintos: I. Pitarch trabaja sobre tablas de datos boléanos y G. Villarroya, sobre matrices gráficas del software TGINF. La tesis de G. Villarroya, se basa en el nuevo programa informático TGINF (Tratamiento Gráfico de la Información) diseñado como resultado del proyecto Bancaixa-2002/05, y que permite visualizar, utilizando la semiología gráfica, los resultados de la clasificación obtenidos, debe permitirnos construir, experimentar y observar situaciones didácticas (lecciones) en el marco curricular de la ESO, que den sentido a la utilización del tratamiento gráfico de datos en esta etapa educativa. En TGINF, se han utilizado las propiedades de representación que permite la semiología gráfica, ya que los resultados gráficos son un instrumento fundamental para memorizar los resultados obtenidos (Bertín; 1977,1988). Para Bertín las significaciones de una construcción gráfica se derivan del juego de oposiciones visuales que aparecen entre los propios elementos

a los que recurre la sintaxis gráfica, y de la relación lógica que se establece visualmente. Estas propiedades nos debe permitir seguir trabajando conjuntamente en las situaciones a construir las nociones de tratamiento y clasificación de datos con el trabajo lógico, incorporando las ventajas que supone la visualización de resultados. Las situaciones didácticas deben incluir el momento de exploración manual de los datos (con permutaciones de filas y de columnas); opción manual que permite TGINF y que permite a los alumnos (a los usuarios, en general) proponer sus propias clasificaciones, con las agrupaciones gráficas que ellos obtienen con las permutaciones de los datos iniciales. La doble posibilidad que ofrece TGINF de mostrar gráficamente de forma automática las clases de datos obtenidas, bien a partir del índice de proximidad de Lerman y de Gras (el mismo que CHIC implementa en forma de grafos y árboles de clasificación), bien a partir del método gráfico propuesto por Bertín (1977,1988), puede servir al profesor (y a los alumnos) como instrumento de comparación entre clasificaciones y de su adecuación en las situaciones propuestas.

Línea de investigación 2. *Contrato didáctico en el proceso de estudio de las matemáticas*

La “Teoría de Situaciones” de G. Brousseau (1986), nos ha permitido avanzar en esta línea de investigación, proporcionándonos los modelos y métodos de análisis y diagnóstico de algunas dificultades inherentes al estudio de las Matemáticas, particularmente la noción de contrato didáctico, así como la posibilidad de solución desde las propias matemáticas, mediante la noción de ingeniería didáctica. Orús (1988, 2002) puso de manifiesto esta dificultad intrínseca a la enseñanza misma de las matemáticas y no identificable a un único concepto matemático: la gestión, por parte del profesor, del razonamiento de los alumnos en los procesos de enseñanza/aprendizaje de las Matemáticas. Esta dificultad Brousseau (1988) la había identificado como un “obstáculo cultural”: el pensamiento natural, junto con el dominio del espacio y la capacidad de enumerar los conjuntos, son conocimientos que el alumno necesita para realizar su aprendizaje, pero que no le son enseñados. Este planteamiento hace necesario diferenciar entre la lógica y el pensamiento natural (tanto del alumno como del enseñante) que se utiliza en las matemáticas y representar de la misma manera ambos razonamientos; para ello Orús ha tomado como instrumento didáctico de modelización el análisis tipológico (análisis clasificatorio) de agregación de datos, no como objeto de enseñanza sino como útil didáctico en dicha resolución: mediante la elaboración y experimentación de una ingeniería didáctica (familia de lecciones y de situaciones didácticas) que actúa sobre la práctica del profesor, permitiéndole negociar el estatus del razonamiento natural (RN) del alumno, superando la ruptura del contrato didáctico habitual, que impide al profesor tratar en clase el RN, y proporcionando al enseñante un instrumento de negociación didáctica respecto de dicho razonamiento. En estas lecciones se utilizan las tablas de doble entrada, con datos booleanos, similares a las utilizadas en el análisis tipológico y éstas van a ser el medio de modelizar diferentes tipos de razonamientos (el pensamiento natural, la clasificación, la lógica, los juicios, etc.) con reglas de manipulación diversas cada uno de ellos, pero repre-

sentados todos con las mismas tablas. Esta misma utilización didáctica va a guiar la elaboración de la ingeniería didáctica para la ESO.

Línea de investigación 3. *Análisis estadístico implicativo (ASI) de datos*

Este método de análisis no simétrico de datos permite la extracción y la estructuración del conocimiento en forma de normas y reglas generalizadas a partir de un conjunto de datos que interrelaciona una población de sujetos (u objetos) con un conjunto de variables. Su origen es la modelización estadística de la quasi-implicación: cuando la variable o la conjunción de variables “a” es observada en la población, entonces generalmente la variable “b” lo es también. Las variables en cuestión pueden ser de distintos tipos: binario, modal, numérico, intervalo, difuso, etc.

Contrariamente a los métodos de análisis simétricos basados, por ejemplo, en una distancia o en una correlación, los conjuntos de reglas obtenidas pueden conducir a hipótesis de causalidad. Estos conjuntos se estructuran según diferentes características comunes complementarias (grafo implicativo, jerarquía orientada). La determinación cuantitativa de los sujetos o descriptores responsables de estas estructuras viene dada por su contribución o su tipicidad. Se accede fácilmente a la visualización de los resultados, así como a su interpretación por medio de CHIC (Classification Hiérarchique Implicative et Cohésitive): un programa informático estadístico adaptado por R. GRAS (1992) a partir de los índices de proximidad o distancia de Lerman. (1981). El programa informático CHIC es un programa innovador, resultado de la colaboración entre investigadores en el campo de la Estadística, de la Didáctica de las Matemáticas (Lerman, Gras y colaboradores de l' École Polytechnique de Nantes del IRMAR de la Universidad de Rennes) (Lerman et coll., 1994) y de la informática (H. Ratsimba-Rajohn, R. Couturier, y otros) A partir de los gráficos -diagramas de árbol- que proporciona este programa informático se pueden obtener clases de equivalencia entre las variables que se tratan. Esta partición, representada gráficamente, del conjunto de los datos, permite obtener conclusiones acerca de la población tratada, tanto si se desea a nivel de clasificación por similitud o por la jerarquía según el grado de implicación entre dichas clases. Así mismo el grafo implicativo que se obtiene con CHIC, permite visualizar las relaciones de “cuasi-implicación” (o implicación estadística) que existe entre las variables (los criterios) según los sujetos que los cumplan y/o entre los sujetos (o los objetos) a partir de los criterios que los definen. El análisis estadístico implicativo y su implementación en el programa CHIC, sigue consolidándose como un método de investigación aplicado en muy diversos campos del conocimiento como se puede apreciar en las Actas de los dos coloquios celebrados en Caen (1995 y 2000) y a los que le han seguido los encuentros internacionales de Sao Paulo (2003) y Palermo (2005) consolidándose un grupo internacional específico de investigación sobre el análisis implicativo, el International Group of Implicative Statistic Analysis, al cual participa este grupo: P. Orús fue miembro del comité científico del III Rencontre Internationale de l'ASI (Analyse Statistique Implicative), Palermo (2005) y se presentó una contribución metodológica sobre la consideración de sujetos suplementarios o “ficticios” en los análisis con CHIC,

como se realiza en otras técnicas de análisis multivariante, como por ejemplo en el análisis factorial de correspondencias, y sus posibles limitaciones (Orús, Gregori, 2005).