

Convocatoria de ayudas de Proyectos de Investigación Fundamental no orientada

MEMORIA TÉCNICA PARA PROYECTOS TIPO A o B

1. RESUMEN DE LA PROPUESTA (Debe rellenarse también en inglés)

INVESTIGADOR PRINCIPAL: Wenceslao González Manteiga

TÍTULO DEL PROYECTO: Metodologías y aplicaciones en estadística semiparamétrica, funcional y espacio-temporal.

RESUMEN

(breve y preciso, exponiendo solo los aspectos más relevantes y los objetivos propuestos)

La estadística semiparamétrica, el análisis de datos funcionales, la modelización con efectos aleatorios, la estimación de conjuntos y la estadística espacio temporal constituyen campos de máximo interés en el entorno de la Estadística en la investigación actual a nivel internacional. En el presente proyecto de investigación se pretende realizar nuevos desarrollos metodológicos en cada uno de estos campos y con especial énfasis en las aplicaciones a contextos estratégicos como el medioambiente, las finanzas y la modelización en áreas pequeñas ligada a la Estadística Oficial. Muchos de los nuevos métodos que se desarrollen serán implementados por primera vez o representarán modificaciones de versiones anteriores, ya implementadas en proyectos anteriores, serán validados mediante los resultados teóricos y de simulación y aplicados explícitamente a problemas susceptibles de transferencia tecnológica. En nuestra solicitud se indican como problemas en esta línea, entre otros, los siguientes: Predicción bidimensional de indicadores de Nox y SO2 en la central térmica de As Pontes (ENDESA), los tests de bondad de ajuste para modelos de tipos de interés en el contexto financiero y la modelización en áreas pequeñas para la predicción de indicadores socioeconómicos para el Instituto Gallego de Estadística.

Palabras clave: Modelización con efectos aleatorios, estadística espacio-temporal, datos funcionales, estimación de conjuntos, modelización semiparamétrica, tests de bondad de ajuste, técnicas de aprendizaje, modelos de tipos de interés, observaciones incompletas, predicción medioambiental.

PROJECT TITLE: METHODS AND APPLICATIONS IN SEMIPARAMETRIC, FUNCTIONAL AND SPATIAL-TEMPORAL STATISTICS

SUMMARY

(brief and precise, outlining only the most relevant topics and the proposed objectives)

Semiparametric statistics, functional data analysis, random effects modeling, set estimation and spatial-temporal statistics are relevant topics within statistics research. In this research project, our aim is to obtain new methodological advances in all this fields, with special emphasis in applications to strategic fields, such as environmental science, finance and small area modeling for Official Statistics. Many of the new developed methods will be implemented for the first time, or they may be modifications of previous versions, already obtained in former projects. These methods will be validated by theoretical results and simulation studies, and they will be applied to technology transfer problems. In this proposal, we include as such problems, among others: prediction of bidimensional indicators for NO_x and SO₂ in the power plant of As Pontes (ENDESA); goodness-of-fit tests for interest rate models in finance and the small area inference for social-economical indicators for the Instituto Galego de Estatística.

Key words: random effects modeling, spatial-temporal statistics, functional data, set estimation, semiparametric modeling, goodness-of-fit tests, learning techniques, interest rate models, incomplete observation, environmental prediction.

2. INTRODUCCIÓN

A) ANTECEDENTES DE TIPO GENERAL.

En la presente convocatoria solicitamos la concesión de ayuda para la realización de nuestro proyecto de investigación, de acuerdo con las líneas generales establecidas en la orden ECI/3354/2007 de 16 de noviembre (BOE, 20 de noviembre de 2007) que establece las bases reguladoras de la concesión de ayudas para el Programa Nacional de Proyectos de Investigación Fundamental en el marco del Plan Nacional de I+D+i 2008-2011 y más concretamente en base a la posterior convocatoria de ayudas para la realización de proyectos de investigación (orden de 26 de noviembre de 2007 (BOE, 30 de noviembre de 2007)). Para ello, teniendo en cuenta tanto los criterios de evaluación por parte de la ANEP, como los correspondientes por parte de las comisiones de selección, que se indican en las anteriores ordenes citadas, hemos perseguido tanto en el diseño de nuestro grupo de investigación, como de la propuesta de investigación a desarrollar los siguientes aspectos:

- Que el grupo mantenga y consolide los niveles de excelencia.
- Que la totalidad de los miembros del equipo tengan la dedicación exclusiva (a excepción obvia de los participantes extranjeros).
- Que se siga con el nivel de eficiencia en el logro de resultados correspondientes a proyectos desarrollados previamente y que han sido financiados por el plan nacional.
- Que se consolide el nivel de internacionalización del grupo.
- Que se continúe con el plan de difusión (ya iniciado en el proyecto anterior) de los resultados obtenidos a nivel de publicaciones en revistas de prestigio, difusión en web del grupo de investigación,...etc., y que se amplíe el interés creciente de diversas empresas en nuestros nuevos desarrollos metodológicos y la consiguiente transferencia tecnológica.

Nuestra línea de trabajo se vino desarrollando dentro del anterior programa Nacional de Matemáticas y cubriendo dentro del campo de la Estadística, entre otros, desarrollos de Estadística no paramétrica, modelos de regresión, remuestreo, aprendizaje estadístico con numerosas aplicaciones, ...etc.. Los resultados obtenidos en la tesis doctoral del IP solicitante del proyecto, Wenceslao González Manteiga, fueron el punto de partida de las investigaciones realizadas en años posteriores por un gran grupo de investigadores en este campo (posiblemente más de 30 en la actualidad, con un total de 22 tesis doctorales dirigidas o codirigidas por el IP de este proyecto), que nutren a las tres Universidades Gallegas y empiezan a introducirse en el marco empresarial, especialmente en el ámbito financiero y biomédico. Al mismo tiempo se desarrollaron y se continúan desarrollando colaboraciones con otros grupos de investigadores nacionales (Universidad Carlos III, Universidad Autónoma de Madrid, Universidad de Zaragoza, Universidad de Granada, Universidad Miguel Hernández de Elche, Universidad de Castellón, Universidad de Cantabria,...etc o internacionales, Universidad de Giessen (Alemania), Universidad de Canberra (Australia), London School of Economics, Universidad de Glasgow, Universidad de Gottingen, Universidad de Carolina del Norte (USA), Universidad de Louvain le Neuve, Universidad de Toulouse,...etc.). El IP de este proyecto pudo observar como, en muchas ocasiones, investigadores de reconocido prestigio, como el Prof. Peter Robinson de la London School of Economics, el Prof. Peter Hall de Australia,...etc. elogiaron la formación de nuestros investigadores, licenciados y doctorados en la Universidad de Santiago de Compostela. Trayectoria similar a la seguida por el propio IP.

Desde hace aproximadamente 17 años venimos disfrutando, prácticamente de forma continua, de ayudas concedidas para la investigación dentro del ámbito de la Promoción General del Conocimiento. Al mismo tiempo el número de investigadores a lo largo de este tiempo se incrementó notablemente. La creación en el año 1989 de las universidades de A Coruña y Vigo, con la consiguiente demanda de profesores del área de Estadística e I.O. en dichas universidades, favoreció también este incremento. Esto supuso, que en la convocatoria de 1998 hubiésemos pedido, ante el gran número de potenciales investigadores, ya entonces existentes, un proyecto coordinado con la Universidad de A Coruña, cuyo coordinador general era el prof. Ricardo Cao Abad, discípulo del IP de este proyecto. En dicha ocasión, a pesar del gran número de publicaciones en las revistas de máximo impacto en nuestra área (Technometrics, JASA, Statistics in Medicine,...etc.) que avalaban nuestra trayectoria investigadora, la concesión que se nos hacía de 2.000.000 pts por cada uno de los dos subproyectos (con 21 investigadores en total) resultaba meramente simbólica. Esto hizo que ambos IPs cambiásemos de estrategia en cuanto a la petición de proyectos coordinados y en la convocatoria correspondiente al 2002, afortunadamente, nuestro grupo de investigación liderado por el IP Wenceslao González Manteiga, logró una financiación correspondiente al 100% de la cantidad solicitada. Análogamente el IP Ricardo Cao logró una financiación con similar cobertura para su grupo de la Universidad de la Coruña. Esta financiación permitió por nuestra parte incrementar nuestra capacidad investigadora, darle mas movilidad a nuestros mas jóvenes integrantes del grupo de investigación y una gran densidad de visitas de investigadores de reconocido prestigio con la organización de seminarios internacionales y el congreso internacional: "The ISI International Conference on Environmental Statistics and Health" celebrado en Santiago de Compostela el verano del año 2003. En la pasada convocatoria el grupo logró la última concesión dentro del plan nacional de Matemáticas con un proyecto MTM2005 muy bien valorado. En los dos últimos años el nivel de excelencia del grupo se acentuó con la entrada del mismo en una red internacional financiada por el gobierno belga, con los mejores grupos de Estadística de las diversas Universidades Belgas y con la coparticipación de otras Universidades, como la Grenoble en Francia, la de Leiden en Holanda o la London School of Higiene and Tropical Medicine en el Reino Unido. Así mismo el grupo fue premiado por la Consellería de Educación de la Xunta de Galicia en el ámbito de consolidación de unidades de investigación competitivas, lo que permitió la contratación de dos investigadores, Rosa María Crujeiras y Beatriz Pateiro, que además de la componente investigadora permite reforzar la componente de consulting dentro del grupo de investigación. El grupo también forma parte del llamando Nodo-Cesga, dedicado en gran medida a potenciar la transferencia tecnológica, y que es parte fundamental del programa consolidar IMATH. El IP es también el coordinador de la Red Mathematica Consulting and Computing de Galicia, que se desarrolla de forma paralela al Nodo-Cesga, con la inclusión de grupos de Electrónica y Computación de Galicia.

B) ANTECEDENTES DE TIPO CIENTIFICO.

La teoría de la inferencia no paramétrica y semiparamétrica de curvas ha alcanzado una etapa de madurez que ha permitido su proyección fuera del ámbito de los investigadores especializados en el tema. Campos como los de la Medicina, Biología, Economía,

Ingeniería, Química, Ciencias Medioambientales, etc. hacen uso en la actualidad de las técnicas que se desarrollaron en los últimos años. Los modelos de predicción espacio temporal para el análisis de polución, las técnicas de aprendizaje en problemas de tipo genético, los nuevos métodos de selección de variable en Econometría, el análisis de fiabilidad en las diversas piezas de un sistema electrónico, el reconocimiento de imágenes, el diseño de activos financieros,... etc. son buen ejemplo de la proliferación del uso de los nuevos métodos basados en la estimación-contrastes de curvas notables: función de densidad, función de regresión, función de distribución, función de autorregresión,... Teniendo, por tanto, en cuenta este panorama de la actualidad investigadora en la inferencia sobre curvas, pretendemos dentro de este proyecto, continuar con los desarrollos de los modelos que venimos estudiando y de otros nuevos, como se detallan a continuación, y continuar con la implementación del consiguiente software que los hará potencialmente útiles a los usuarios de los diversos campos, como, por ejemplo, los antes comentados. Nos encontramos por tanto con el reto de seguir organizando, desde un punto de vista computacional, todas las técnicas que se elaboren desde un punto de vista metodológico, contrastando previamente su validez mediante el análisis teórico y los consiguientes estudios de simulación. Para mantener esta finalidad, integran nuestro grupo de investigación: personal investigador del que se dispone en la actualidad en las universidades gallegas en el campo de la inferencia sobre curvas, fundamentalmente de la Universidad de Santiago con algunas personas de la Universidad de Vigo y de la Universidad de la Coruña e investigadores de reconocido prestigio a nivel internacional, con los que mantenemos actualmente una fuerte conexión en el desarrollo investigador. En total 21 investigadores, con personal cercano al mundo de la computación y personal mas próximo a los aspectos más teóricos. En general los miembros nacionales de nuestro grupo actual se distribuyen en diversos centros, tanto del departamento de Estadística e I.O. de la Universidad de Santiago, a donde pertenece el IP, como en otros centros: Facultad de Ciencias Económicas de la Universidad de Santiago y de la Universidad de Vigo, Escuelas de Ingeniería de las Universidades de Vigo y Coruña, etc. y que representan en gran medida el nexo con las diversas ciencias experimentales o tecnológicas. Creemos que todo esto precisamente constituye un gran aval para la consecución de los objetivos propuestos. El grupo de investigación lo hemos seleccionado atendiendo a tres criterios fundamentales. Por un lado el criterio de excelencia con investigadores nacionales consolidados, con gran dedicación en los últimos años, como lo demuestra su curriculum vitae, y con diversas publicaciones de impacto. Por otro lado el criterio de renovación científica con jóvenes investigadores con proyección hacia el futuro, que están iniciando o realizando sus tesis doctorales, algunos como todavía alumnos del programa de doctorado con mención de Calidad del ámbito de la Estadística y que se imparte de forma interuniversitaria en las Universidades Gallegas y que representan en general a estudiantes con un muy buen expediente de licenciatura o doctores que han presentado recientemente su tesis doctoral bajo la dirección del IP. Finalmente atendiendo al criterio de internacionalización de la actividad realizada seguimos en nuestro grupo con dos personas de un gran nivel científico. Los profesores Winfried Stute de la Universidad de Giessen (futuro Honoris Causa por la Universidad de Santiago de Compostela) e Ingrid Van Keilegom del Instituto de Estadística de Louvain le Neuve, con los que venimos desarrollando una gran labor investigadora conjunta. Ambos profesores son autoridades a nivel mundial en el campo de la Inferencia no Paramétrica. El prof. Stute, con el que ya colaboramos desde hace muchos años y la profesora Van Keilegom con la que venimos colaborando desde el proyecto anterior con trabajos de investigación conjuntos. Así mismo en esta ocasión se incorpora al grupo el profesor Stefan Sperlich, antiguo profesor de la Universidad Carlos III y ahora full professor de la Universidad de Göttingen.

En lo que sigue nos centraremos en describir los antecedentes en los diversos objetivos concretos desde un punto de vista metodológico y poniendo de relieve si se trata de un objetivo totalmente nuevo o una prolongación en base a nuevos estudios dentro de temáticas ya existentes en proyectos anteriores. Muchas de las referencias (las que forman parte de la lista de aportaciones por parte de miembros del grupo en revistas del Journal Citation Reports en los últimos años) están indicadas en la sección 6. Del resto, por razones de limitación de espacio, en el propio texto se indican datos de las mismas.

1) Modelización con efectos aleatorios.

Una metodología que ha adquirido especial importancia en los últimos años es la modelización con efectos mixtos (efectos fijos y efectos aleatorios). Los modelos mixtos son una extensión de los modelos de regresión que asumen coeficientes de regresión fijos y aleatorios. Estos modelos son importantes cuando las observaciones están correladas. Por ejemplo, las medidas repetidas son recogidas para el mismo individuo en distintos instantes de tiempo. Lo que nos lleva a considerar que las observaciones para el mismo individuo están correladas, especialmente si son recogidas en un corto período de tiempo. Esto hace que sean muy populares en muchas áreas de la estadística, análisis de datos panel (véase el libro de Diggle, Heagerty, Liang and Zeger; 2002, Oxford University Press), datos cluster, inferencia en áreas pequeñas (véase el libro de Rao, 2006, Wiley), etc. Para una revisión y aplicaciones de estos modelos se puede consultar el libro de J.Jiang (2007, Springer).

Recientemente, se ha introducido el mundo de la no paramétrica, permitiendo avanzar hacia modelos de efectos mixtos semiparamétricos generalizados. Podemos citar el artículo de Opsomer, Claeskens, Ranalli, Kauermann y Breidt (JRSS B, por aparecer), que introducen los splines en la estimación de estos modelos, con aplicación a la problemática de áreas pequeñas; Lombardía y Sperlich (FUNCAS, 2006) introducen un procedimiento de estimación para estos modelos con funciones tipo núcleo y un test de especificación de modelos semiparamétricos con bootstrap, ilustrándolo con una aplicación en áreas pequeñas.

Finalmente, debemos resaltar la aplicación de estos modelos a problemas reales en distintos contextos como biomedicina, agricultura, economía y estudios de ciencias sociales.

2) Inferencia en modelos de estadística espacial y espacio temporal.

En el análisis de datos espaciales o espacio-temporales, las dos componentes básicas y que describen los dos tipos de variabilidad son la tendencia y la estructura de dependencia. Aunque existe una amplia literatura relativa a métodos de estimación y predicción en modelos espaciales (Cressie, Wiley-1993) no existen apenas referencias en lo relativo a técnicas de contraste de hipótesis.

El problema del modelado y contraste de la estructura de dependencia de datos espaciales a través de técnicas espectrales se aborda en la tesis doctoral de Rosa Crujeiras (2007). En este contexto, se proponen contrastes de bondad de ajuste y un contraste para la comparación de densidades espectrales (Crujeiras y otros, 2007). Aunque el enfoque espectral técnicamente atractivo, desde el punto de vista práctico presenta algunas limitaciones ya que uno de los requerimientos básicos de las técnicas espectrales es que los datos de las

muestras estén regularmente espaciados. Siguiendo en el dominio espectral, es interesante la extensión de las técnicas estudiadas al caso de datos irregularmente espaciados (en espacio o en espacio y tiempo), a través de modificaciones de la transformada de Fourier.

A pesar de las ventajas metodológicas del dominio espectral, en la aplicación práctica a otras áreas, los científicos prefieren estudiar /modelar la dependencia a través de herramientas como el variograma, cuya interpretación resulta más simple al estar planteado en el dominio espacial. Es por tanto interesante el plantear contrastes utilizando el variograma, a través de técnicas no paramétricas y funcionales. Diblasi y Bowman (Biometrics, 2001) y Diblasi y Maglione (SERRA, 2004) plantean un contraste de independencia y de bondad de ajuste (para hipótesis nula simple), respectivamente, basados en un estimador no paramétrico del variograma. Desde esta perspectiva, se puede abordar el contraste de isotropía, de bondad de ajuste y de estacionariedad (hipótesis básicas en el modelado de datos espaciales). Hemos establecido contactos con el Prof. Bowman, con el que estamos colaborando para el desarrollo de nuevos contrastes basados en estimaciones del variograma, utilizando técnicas de binning para su aplicación práctica. También el planteamiento de contrastes desde el punto de vista funcional puede resultar de utilidad.

Otro de los campos en los que se ha venido trabajando es el desarrollo de metodología de estimación y contrastes para procesos puntuales, de potencial aplicación a datos forestales y que constituye el tema de tesis de Isabel Fuentes.

La variabilidad a gran escala en el contexto de datos espaciales, se modela a través de modelos de regresión espacial y espacio-temporal (véase, por ejemplo, Holland y otros (Environmetrics, 2000)), donde la variabilidad a pequeña escala –es decir, la estructura de dependencia- se considera un nuisance parameter. Un objetivo, de máximo interés actual es el desarrollo de contrastes de bondad de ajuste sobre la tendencia y sobre la estructura de correlación en un modelo de estadística espacial, problema que se puede abordar haciendo uso de los procesos empíricos de regresión.

3) Modelización con datos funcionales.

La estadística con datos funcionales ha experimentado un desarrollo notable en los últimos tiempos en paralelo a la aplicación de la computación para procedimientos de medida. Así, en multitud de áreas se ha empezado a trabajar, cada vez con mayor frecuencia, con observaciones de una variable aleatoria tomadas a lo largo de un intervalo continuo. Por ejemplo, para el control medioambiental es frecuente contar con curvas de mediciones en tiempo real de posibles contaminantes, en Economía tenemos las curvas intra-día de cotizaciones que se recogen instantáneamente o en Ingeniería tenemos curvas de producción o demanda eléctrica. Normalmente, estos datos son almacenados en grandes bases de datos aunque el objetivo perseguido con el análisis estadístico no difiere del clásico: explorar y describir el conjunto de datos resaltando características importantes, explicar y modelar la relación entre una o varias variables (ambas de tipo funcional o no), discriminar grupos (conocidos o no) de un conjunto de datos, contrastar hipótesis, validarlas y predecir alguna característica para una nueva observación. Los autores que más han contribuido a la popularización de las técnicas de datos funcionales son sin duda alguna Ramsay y Silverman que han publicado dos libros (1997, 2002) fundamentalmente dedicados a la aplicación a problemas notables. El otro hito bibliográfico importante de los últimos años es el libro de Ferraty y Vieu (2006) que establece ya marcos teóricos para el desarrollo de nuevos procedimientos. Del carácter emergente de la disciplina da idea los números monográficos especiales que varias revistas han publicado recientemente (Statistica Sinica, Vol. 14, 2004; Computational Statistics & Data Analysis, Vol. 51, 2007; Computational Statistics, Vol. 22, 2007) o que si contamos con scholar.google.com el número de trabajos científicos que incluyen “functional data analysis” en su texto, aparecen más de 900 entradas desde el 2003.

El análisis de estos datos por su propia naturaleza entraña importantes dificultades en todos los ámbitos. Así, en el análisis descriptivo de datos funcionales se ha avanzado mucho en los últimos años donde podemos destacar los trabajos de Dabo-Niang et al (Computational Statistics & Data Analysis, 2007), Cuevas y otros (2004, 2007) que presentan alternativas a los estimadores clásicos de localización y escala. También se ha avanzado en el objetivo de identificar datos anómalos como en los trabajos de Febrero y otros (2007, 2008) de vital importancia para el tratamiento de datos reales. En el apartado de relacionar variables (regresión funcional) hay más camino andado. Aquí el caso más estudiado es el de variable respuesta escalar y variable regresora funcional bajo diseño aleatorio (véase los trabajos de Cardot y otros (Statistics & Probability Letters, 1999; Scandinavian Journal of Statistics, 2003) el libro de Ramsay y Silverman o el libro de Ferraty y Vieu en un contexto no paramétrico). Bajo diseño fijo es destacable el trabajo de Cuevas y otros (Canadian Journal of Statistics, 2002) donde ambas variables son funcionales. Considerando el ANOVA como una extensión de un modelo de regresión con respuesta funcional y variable regresora discreta, el trabajo de Cuevas y otros (2004) diseña un contraste para encontrar diferencias entre grupos en datos funcionales. El problema de la clasificación (supervisada o no) ha sido también extensamente estudiado. Citas relevantes en este apartado son el propio libro de Ferraty y Vieu antes mencionado o los trabajos de James y Sugar (JASA, 2003), Tarpey y Kinateder (Journal of Classification, 2003) o Yao y otros (JASA, 2005). Como una extensión robusta del algoritmo de *k*-medias se presenta el trabajo de Cuesta-Albertos y Fraiman (Computational Statistics & Data Analysis, 2007). Sin embargo, en mayor o menor medida, los trabajos citados se han dedicado más a la estimación que al contraste. Un trabajo reciente que posiblemente cambiará este planteamiento es el de Cuesta-Albertos y otros (Journal of Theoretical Probability, 2007) que permite caracterizar poblaciones funcionales mediante proyecciones aleatorias y que ya está permitiendo el desarrollo de nuevas técnicas para datos funcionales con tratamientos bastante simples. Esto va a permitir satisfacer la demanda de métodos estadísticos funcionales que áreas como medioambiente, industria o economía reclaman.

4) Estimación de conjuntos.

La estimación de conjuntos aborda el problema de recuperar un conjunto a partir de la observación de una muestra aleatoria de puntos cuya distribución está relacionada con él de forma más o menos directa. Por ejemplo, los datos pueden ser una muestra aleatoria simple de una distribución *F* y el conjunto de interés puede ser el soporte de probabilidad de *F*. Las técnicas de estimación de conjuntos se han utilizado, entre otras áreas, en el análisis de conglomerados, el análisis de imágenes o la economía. En Cuevas y Rodríguez Casal (2003) (capítulo del libro Recent Advances and Trends in Nonparametric Statistics, North Holland, 2003) se recogen algunas de estas aplicaciones.

El interés en esta línea de trabajo surgió de la gran relación existente con el grupo de investigación dirigido por el profesor Antonio Cuevas, de la Universidad Autónoma de Madrid. Fruto de esa colaboración es la tesis doctoral de Alberto Rodríguez Casal (Santiago, 2003) codirigida por los profesores Antonio Cuevas y Wenceslao González Manteiga. En esa tesis doctoral se aborda el problema de la

estimación de conjuntos desde un punto de vista geométrico. La memoria se centra en la estimación de la frontera, característica esencial cuando se desea recuperar la forma de un conjunto. Así, se propone un nuevo criterio, basado en la calidad de la reconstrucción de la frontera, para evaluar la eficacia de los estimadores (Cuevas y Rodríguez-Casal 2004) y se analizan algunos estimadores existentes en la literatura empleando este nuevo criterio. También se proponen nuevos estimadores del soporte de una distribución de probabilidad que verifica ciertas restricciones de forma de tipo convexo (Rodríguez-Casal 2007) y se aborda el problema de la estimación de conjuntos de nivel de funciones generales no necesariamente definidas en el espacio euclídeo (Cuevas, González-Manteiga y Rodríguez-Casal 2006). Las técnicas propuestas en este último trabajo se aplican al análisis cluster con datos esféricos.

La colaboración con el profesor Antonio Cuevas en estimación de conjuntos ha continuado a lo largo de los últimos años. Fruto de este trabajo es el artículo recientemente publicado en *Annals of Statistics* (Cuevas, Fraiman y Rodríguez-Casal 2007) donde se propone un nuevo estimador, universalmente consistente, de la longitud de la frontera de un conjunto. El cociente entre la longitud de la frontera y la raíz cuadrada del área del conjunto proporciona un índice, denominado Índice de Contorno, que es invariante por cambios de escala y que permite cuantificar el nivel de fragmentación de un conjunto. Este índice ha sido utilizado en campos como la cardiología o la oncología para evaluar el efecto de diversos tratamientos. El cálculo del Índice de Contorno requiere disponer de una buena aproximación de la longitud de la frontera, cuestión no totalmente trivial cuando existe ruido aleatorio en la imagen disponible. El problema fundamental del estimador propuesto en Cuevas, Fraiman y Rodríguez-Casal (2007) es su lenta convergencia cuando el conjunto de interés es suave, ya que no presupone ninguna condición de forma especial (lo que garantiza a su vez su carácter universal). Si se presupone que el conjunto a estimar verifica una condición de forma (no demasiado restrictiva) denominada alfa-convexidad (véase el artículo en *Annals of Statistics* de Walther 1997) entonces es posible proponer un estimador de la longitud de la frontera más eficiente, tal como se demuestra en Pateiro-López y Rodríguez-Casal (2007), trabajo que actualmente se encuentra en segunda revisión en *Advances in Applied Probability*.

5) Modelos no paramétricos y semiparamétricos.

1) Modelización semiparamétrica en predicción medioambiental.

Como consecuencia de la colaboración de nuestro grupo de investigación con la Unidad de Producción Térmica de As Pontes, propiedad de Endesa Generación S.A., se ha implantado un sistema de control suplementario de la contaminación atmosférica en dicha central. Dicho sistema ha seguido diversas evoluciones a lo largo de los años con diversos modelos de predicción de tipo semiparamétrico del SO₂ que se han incorporado a la utilidad correspondiente, a destacar:

- 1) Un modelo de predicción no paramétrico con errores ARI, García Jurado y otros (Technometrics, 1995)
- 2) Modelización de tipo semiparamétrico espacio-temporal, Angulo y otros (Environmental and Ecological Statistics, 1998)
- 3) Modelización con modelos parcialmente lineales, Prada Sánchez y otros (Environmetrics, 2000)
- 4) Modelización usando redes neuronales, Fernández de Castro y otros (Air & Waste Manage, 2003)
- 5) Predicción con modelos de respuesta binaria de tipo semiparamétrico, Roca Pardiñas y otros (Environmetrics, 2004)
- 6) Predicción con repuestas de tipo funcional, Fernández de Castro y otros (Technometrics, 2005)
- 7) Estimación y contrastes con modelos aditivos generalizados, Roca Pardiñas y otros (Statistics and Computing, 2005)
- 8) Predicción usando técnicas de aprendizaje tipo boosting, Fernández de Castro y González Manteiga (Stoch. Environ. Res. Risk. Asses., 2007)

En la actualidad vivimos un momento apasionante desde el punto de vista científico. La Central Térmica está sufriendo un proceso de adaptación que permitirá su funcionamiento con 100% de carbón de importación, las emisiones de SO₂ se reducirán un 95% y empezará a funcionar un nuevo ciclo combinado en donde la serie de NO_x cobra vital importancia. Estamos ante el desafío de predecir un serie con respuesta bidimensional en donde los errores bidimensionales pueden tener distinta estructura paramétrica y tengamos por tanto que jugar con aspectos de cointegración.

2) Tests de bondad de ajuste.

El desarrollo en el campo de los tests de bondad de ajuste para modelos de regresión ha sido muy intenso en los últimos 18 años. Por brevedad, nos limitaremos a describir solamente lo realizado por nuestro grupo. Así, mediante el uso de métodos de suavización, se han establecido procedimientos de contraste de bondad de ajuste y de igualdad de funciones de regresión, obteniéndose su distribución asintótica cuando los errores del modelo son dependientes (González Manteiga y Vilar Fernández (CSDA,1995), Vilar Fernández y González Manteiga (Statistics,2004)). Además, se ha introducido un mecanismo de remuestreo bootstrap, probando su buen funcionamiento para el problema de bondad de ajuste anteriormente citado (Vilar Fernández y González Manteiga (Commun. in Stat.,1996), (CSDA, 2000)). También se estudió la posibilidad de usar pilotos no paramétricos del estilo regresión polinómica local, como se puede ver en Alcalá y otros (Stat. Prob. and Letters, 1999), de tipo núcleo adaptado a la regresión con respuesta binaria (JSPI, Rodríguez Campos y otros (JSPI, 1998)) o de tipo k-puntos próximos, como se puede ver en Stute y González Manteiga (JSPI, 1996). Muchos de estos procedimientos se extendieron a datos dependientes (González Manteiga y otros (2002)). También se ha estudiado la utilización del método bootstrap para la bondad de ajuste en la regresión (Stute y otros (JASA, 1998)) ,(Stute y otros (Commun in Stat., 2000) y (Delgado y González Manteiga (2001))), obteniéndose un muy buen comportamiento de los métodos propuestos.

Se ha desarrollado una amplia teoría para los tests de bondad de ajuste, de tipo ji cuadrado, bootstrap o asintótico normal, basados en distintas formas de suavización: desarrollos ortogonales, tipo núcleo, spline, k puntos próximos, como se puede ver en el trabajo de Ramil y González Manteiga (Commun. in Stat, 1998). Este trabajo se complementa con el análisis de la varianza no paramétrico de Ramil y González Manteiga (Stat. Sinica,2000). Aquí se desarrollan las tablas del análisis de la varianza en un contexto no paramétrico, que incluyen la nueva componente del sesgo, inexistente en los contextos tradicionales. Mas recientemente en una colaboración del IP con Ingrid Van Keilegom que actuó como codirectora de la tesis doctoral, modalidad europea, del estudiante Juan Carlos Pardo, se desarrollo una amplia teoría de tests de bondad de ajuste para modelos de regresión basada en la distribución de los residuos, ver Pardo y otros (2006 y 2007). Ver también Sello y otros (2007).

3) Modelización con datos sesgados, censurados y/o truncados y faltantes.

En los últimos años el grupo trabajó en diversas líneas en las que se persigue desarrollar metodología adaptada a situaciones en las que la muestra inicial es incompleta en algún sentido. A destacar:

- 1) Adaptado al contexto del Análisis de Supervivencia con censura proporcional se desarrolló una teoría de tests de bondad de ajuste de tipo ji-cuadrado para modelos de distribución en (de Uña Álvarez y otros (2003)).
- 2) Con el uso de integrales Kaplan Meier se desarrolló una amplia teoría para la bondad de ajuste de modelos de regresión en los que la variable respuesta puede estar censurada y/o truncada (Sánchez Sellero y otros (2005)).
- 3) La Colaboración del IP con Ana Pérez González dio lugar a la tesis de esta última con varias publicaciones dedicadas a modelos de regresión en los que respuesta puede ser faltante (González Manteiga y Pérez González (2004) (2006), González Manteiga y otros (2004)).
- 4) Por último, recientemente el IP inició una nueva línea de trabajo para los tests de bondad de ajuste para modelos de regresión, en base a procesos empíricos, en los que la respuesta puede estar sesgada (por ejemplo longitudinalmente). Este estudio se está llevando a cabo con José Antonio Cristóbal y Jorge Ojeda de la Universidad de Zaragoza. EL grupo de Zaragoza al que pertenecen estos investigadores ya tiene experiencia de publicaciones previas de tests de bondad de ajuste con datos sesgados pero utilizando técnicas de suavización

4) Modelos parcialmente lineales.

En los últimos años también hemos llevado a cabo un amplio estudio para contrastes acerca de las diversas componentes paramétricas y no paramétricas de un modelo parcialmente lineal, en una primera fase para modelos con estructura de dependencia en los errores del tipo memoria corta (González Manteiga y Aneiros Pérez (2004)) y posteriormente para modelos con errores del tipo memoria larga (Aneiros Pérez y otros (2004)). Últimamente hemos trabajado en aplicaciones al contexto financiero con Juan Carlos Reboredo, experto en fundamentos del análisis económico, e investigador que afortunadamente se incorpora a nuestro grupo en esta ocasión y fortalece notablemente la capacidad del grupo en las aplicaciones económicas. En particular estamos trabajando en modelos con estructura FARIMA-GARCH en los errores.

Al mismo tiempo en los últimos años hemos iniciado una nueva línea de trabajo con Graciela Boente y su grupo de la Universidad de Buenos Aires con el intento de desarrollar metodología en modelos parcialmente lineales robustificados y que puedan tener observaciones faltantes en la respuesta. Un primera aproximación a este estudio se presentó en una sesión invitada del último congreso del ISI celebrado en Lisboa en el 2007.

5) Modelos en técnicas de aprendizaje

En los últimos años y especialmente desde la aparición de textos como por ejemplo el libro de Hastie, Tibshirani y Freedman (2001), las técnicas de aprendizaje se hicieron muy populares en el mundo de la Estadística (Statistical Learning) y también en el de las Ciencias de la Computación (Machine Learning). Dentro de las diversas técnicas que se ofrecen en el mercado cabe destacar las basadas en redes neuronales y las llamadas SVM (Support Vector Machine). En nuestro grupo de investigación se hicieron diversos desarrollos de tipo metodológico o de aplicación, entre los que podríamos destacar:

- 1) Conexión de las técnicas de SVM con el kriging en Estadística Espacial (Matías y González Manteiga (2006))
- 2) Conexión de las técnicas SVM con los modelos parcialmente lineales (Matías (2005), ver CV de Matías)
- 3) Aplicaciones de las SVM al contexto financiero y a la prospección minera entre otras (Matías y otros (2004), Matías y otros (2008))
- 4) Modelización con redes neuronales y aplicaciones (Matías y otros (2004), Ordoñez y otros (2005), Taboada y otros (2006, 2007), Matías y otros (2007, 2008), ver para mas detalles CV de Matías)

Muy recientemente algunos miembros del grupo empezamos a bordar la posibilidad de estudios de conexión entre las técnicas de aprendizaje y los datos de tipo funcional.

6) Modelización financiera de series de tipo de interés

El análisis de series de tiempo, tanto desde el punto de vista teórico como empírico, constituye una parte fundamental en el estudio de los mercados financieros. Una de las principales cuestiones de la economía financiera es la predictibilidad del precio de activos.

Dado el carácter aleatorio de los precios, se han desarrollado diversas técnicas estadísticas, tanto de tipo paramétrico como no paramétrico, para la predicción de valores futuros. Dentro de la clase de modelos de series de tiempo lineales se incluyen los clásicos modelos ARMA. Estos constituyen una de las familias de modelos paramétricos más utilizada en el análisis de series de tiempo. (Brockwell y Davis. (Springer, 1987)). Sin embargo, a la vista de ciertas limitaciones de estos modelos, principalmente cuando el objetivo es modelizar la varianza condicional o volatilidad, se han introducido en los últimos años modelos de series de tiempo no lineales, como se puede ver en Fan y Yao (libro de Springer, 2003). Dentro de este grupo se incluirían por ejemplo los modelos de tipo ARCH y GARCH y ramificaciones de estos. La motivación principal de estos modelos es evaluar y predecir el riesgo en series de tiempo en finanzas. Podemos destacar también el desarrollo reciente de modelos basados en redes neuronales, ver Tsay (libro de Wiley, 2005).

De especial relevancia, por las implicaciones en la economía directa familiar, la interbancaria, etc son las series de tipo de interés. En los últimos 25 años para la modelización en tiempo continuo de las mismas se utilizaron diversos modelos de difusión. Por ejemplo el GBM entre otros muchos. Una amplia familia de modelos viene dada por los de tipo homogéneo, ver por ejemplo el artículo de Fan y Zhang (JASA, 2003), en donde los parámetros caracterizantes del modelo no cambian con el tiempo.

La literatura también está repleta de modelos de tipo no homogéneo. Por ejemplo extensiones de tipo no homogéneo de famosos modelos, como el CKLS, estudiadas por Fan, Jiang y Zhou (Stat. Sinica, 2003). Ver también el reciente paper de Fan y Jiang, publicado como artículo invitado en TEST, 2007. Una versión discretizada de este modelo se puede reescribir como un modelo de regresión. Tomando como base la modelización con un modelo de regresión, son de especial relevancia en la actualidad los llamados "test de bondad de ajuste" para modelos de tipos de interés. Partiendo de la generalidad de los modelos no homogéneos antes descritos. Fan y sus coautores desarrollaron mecanismos muy recientes de bondad de ajuste basados en ideas ligadas a la razón de verosimilitudes y técnicas de suavización en donde la hipótesis nula representa el modelo de partida para la serie de tipos de interés y la alternativa puede venir dada por un modelo genérico de tipo no homogéneo. De nuevo se recomienda la lectura del reciente artículo invitado publicado en TEST, 2007.

3. OBJETIVOS DEL PROYECTO

(máximo 2 páginas)

3.1. Describir brevemente las razones por las cuales se considera pertinente plantear esta investigación y, en su caso, la **hipótesis de partida** en la que se sustentan los objetivos del proyecto

(máximo 20 líneas)

Según se desprende del apartado anterior, la modelización no paramétrica, semiparamétrica y con datos funcionales es actualmente una parte representativa de la investigación en inferencia estadística. Con la potencialidad de numerosos desarrollos novedosos, como por ejemplo el estudio con datos funcionales, la inferencia bootstrap adaptada a este contexto, el estudio de datos espaciales, espacio temporales y de procesos puntuales, la creación de nuevos modelos con componentes no paramétricas para contextos notables, como por ejemplo los modelos de tipos de interés en finanzas o el estudio de los modelos con efectos aleatorios, la generalización de los problemas de predicción ambiental, la estimación de conjuntos,...etc. consideramos que las ideas, de carácter teórico y/o aplicado, que proponemos estudiar son sin duda un estímulo para seguir investigando en esta temática. Fruto del gran dinamismo investigador en estos campos, por parte de nuestro grupo, son sin duda las diversas tesis presentadas recientemente, todas ellas avaladas con publicaciones internacionales y desarrolladas con la modalidad de doctorado europeo. Otra razón de muchísimo peso para el desarrollo de nuestros objetivos es la demanda por parte de diversas empresas de metodología en la línea de nuestro trabajo adaptada a sus necesidades, a destacar, la predicción de la renta mixta a nivel comarcal por parte del Instituto gallego de Estadística (con técnicas de inferencia en áreas pequeñas), la predicción de niveles de SO₂ y NO_x en torno a la central térmica de As Pontes, la modelización financiera para Caixa Galicia (con técnicas semiparamétricas) o las aplicaciones de las técnicas de datos funcionales a predicción medioambiental o para la industria cerámica.

3.2. Indicar los **antecedentes y resultados previos**, del equipo solicitante o de otros, que avalan la validez de la hipótesis de partida

Creemos que el interés del tema está suficientemente avalado por su riqueza metodológica y las numerosas posibles aplicaciones. En lo que respecta a los resultados previos del equipo, se puede señalar que en el período 2001-2007 los integrantes de este grupo han publicado más de 100 artículos en revistas internacionales incluidas en el Journal Citation Reports. En el apartado 6 de esta memoria se pueden ver más detalles sobre esto. En los impresos ligados a la solicitud tipo C está la lista completa.

3.3. Enumerar brevemente, pero con claridad, precisión y de manera realista (es decir, acorde con la duración prevista del proyecto) los **objetivos concretos** que se persiguen. La novedad y relevancia de los objetivos (así como la precisión en la definición de los mismos) se mencionan explícitamente en los criterios de evaluación de las solicitudes

1. Se indican a continuación las principales líneas de investigación que proponemos seguir. En el apartado siguiente de esta memoria (Metodología y plan de trabajo) se dan algunos detalles adicionales y se indica la distribución de tareas.

1. Modelización con efectos aleatorios.

a) Modelos mixtos.

En esta línea de investigación se está colaborando con el Prof. D. Stefan Sperlich de la Universidad de Göttingen (Alemania) y D^a María Dolores Martínez Miranda de la Universidad de Granada. Los objetivos marcados son los siguientes:

- Estudio y desarrollo de métodos de estimación no paramétrica para modelos mixtos con aplicaciones.
- Estudio del parámetro ventana para la estimación tipo núcleo de modelos mixtos.
- Estudio y desarrollo de contrastes para modelos mixtos.
- Diseño de nuevos mecanismos de remuestreo para este contexto.
- Aplicación de la metodología desarrollada a bases de datos de interés.

b) Modelos mixtos en áreas pequeñas.

En esta línea de investigación se están realizando estudios conjuntamente con el Prof. D. Morales de la Universidad Miguel Hernández de Elche. Los objetivos marcados son los siguientes

- Estimación no paramétrica del Error Cuadrático Medio de algunas características en áreas pequeñas (medias, totales, etc.) y sus posibles estimadores, para modelos mixtos con un factor aleatorio.
- Estudio de contrastes de significación del factor aleatorio en modelos mixtos utilizados en la estimación en áreas pequeñas.
- Diseño de nuevos mecanismos de remuestreo para este contexto.
- Elaboración de software y aplicación de la metodología desarrollada a bases de datos de interés (Por ejemplo: estimación de la Renta media en las comarcas de la Comunidad Autónoma de Galicia)

2. Inferencia en modelos de estadística espacial y espacio temporal.

- Extensión de las técnicas de estimación y contrastes en el dominio espectral a datos irregularmente espaciados.
- Desarrollo de métodos de contraste para modelos de estadística espacial y espacio temporal basados en estimaciones no paramétricas y funcionales del variograma.
- Desarrollo de contrastes para la tendencia de datos espaciales, utilizando procesos empíricos de regresión.
- Desarrollo de técnicas de estimación y contrastes no paramétricos para procesos puntuales.
- Desarrollo de software para la aplicación de la metodología desarrollada a datos medioambientales, ecológicos y forestales.

3. Modelización con datos funcionales.

- Desarrollo de métodos de contraste para datos funcionales basados en proyecciones aleatorias.
- Desarrollo de estimadores para modelos lineales funcionales con covariables funcionales.
- Desarrollo y extensión de metodologías de contraste para modelos de regresión con datos funcionales.
- Estudio y desarrollo de contrastes para distintas modelizaciones de datos funcionales basados en remuestreo.
- Desarrollo de software y aplicación a datos de tipo medioambiental o industrial

4. Estimación de Conjuntos.

- Estimación de la longitud de la frontera
- Estimación del soporte de una distribución de probabilidad
- Contraste de hipótesis sobre la forma del soporte de una distribución de probabilidad
- Aplicaciones de la estimación de conjuntos a problemas de interés: clasificación, análisis de imágenes, etc.

5. Modelos no paramétricos y semiparamétricos.

a) Modelización semiparamétrica en predicción medioambiental.

- Generalización a modelos semiparamétricos con respuesta multidimensional.
- Estimación, contrastes y predicción en dichos modelos.
- Aplicación al sistema de predicción en la Central Térmica de As Pontes para las variables SO₂ y NO_x.

b) Tests de bondad de ajuste.

- Nuevos avances para tests basados en procesos empíricos marcados por las covariables.
- Robustificación de los tests de bondad de ajuste.
- Tests basados en verosimilitud empírica.

c) Modelización con datos sesgados, censurados y/o truncados y faltantes.

- Contrastes con procesos empíricos para modelos de regresión con respuesta censurada.
- Contrastes con procesos empíricos para modelos de regresión con respuesta faltante.
- Contrastes con procesos empíricos para modelos de regresión con respuesta sesgada.
- Estimación y contrastes para curvas ROC, datos completos, censurados, etc.

d) Modelos parcialmente lineales.

- Inferencia en modelos parcialmente lineales robustos y con datos faltantes.
- Contrastes en modelos parcialmente lineales con errores FARIMA-GARCH. Aplicaciones a la economía.

e) Modelos en técnicas de aprendizaje.

- Nuevos avances para clasificación con SVM.
- SVM con datos funcionales.
- Aplicaciones al entorno industrial, financiero y genómico.

6. Modelización financiera en series de tipos de interés.

- Selección, alternativa a las basadas en técnicas de suavización, de modelos usando procesos empíricos de regresión.
- Aplicaciones a series de tipo de interés a diversos plazos.
- Aplicaciones a bases de datos suministradas por la fundación Caixa Galicia.

3.4. En el caso de Proyectos Coordinados, el coordinador deberá indicar (máximo dos páginas):

- los objetivos globales del proyecto coordinado, la necesidad de dicha coordinación y el valor añadido que se espera alcanzar con la misma;
- los objetivos específicos de cada subproyecto;
- la interacción entre los distintos objetivos, actividades y subproyectos;
- los mecanismos de coordinación previstos para la eficaz ejecución del proyecto.

4. METODOLOGÍA Y PLAN DE TRABAJO

(en el caso de proyectos coordinados deberá abarcar a todos los subproyectos)

Se deben **detallar y justificar con precisión la metodología y el plan de trabajo** que se proponen y debe exponerse la planificación temporal de las actividades, incluyendo cronograma (se adjunta un posible modelo).

- El plan de trabajo debe desglosarse en actividades o tareas, fijando los hitos que se prevé alcanzar en cada una de ellas. En los proyectos que empleen el Hespérides o se desarrollen en la zona antártica, deberán también incluir el plan de campaña en su correspondiente impreso normalizado.
- En cada una de las tareas, deben indicarse el centro ejecutor y las personas (ver apartados 2.1, 2.2 y 2.3 del formulario de solicitud) involucradas en la misma. Si en el proyecto participan investigadores de otras entidades no relacionados en el apartado 2.3 del formulario de solicitud, deberán exponerse los méritos científicos que avalan su participación en el proyecto.
- Si solicita ayuda para personal contratado, justifique claramente su necesidad y las tareas que vaya a desarrollar. Recuerde que sólo podrá solicitar costes de personal en régimen de contratación, **no se podrán asignar becarios con cargo al capítulo de personal** del proyecto.

La adecuación de la metodología, diseño de la investigación y plan de trabajo en relación con los objetivos del proyecto se mencionan explícitamente en los criterios de evaluación de las solicitudes.

Teniendo en cuenta a los diversos investigadores que constituyen el grupo de investigación:

Investigadores nacionales:

-**Wenceslao González Manteiga** (Profesor de la USC, IP del proyecto, colaborador en la mayoría de los campos de investigación del proyecto)

-**Manuel Febrero Bande** (Profesor de la USC, experto en estadística computacional, estadística espacial y datos funcionales)

-**Jose Manuel Prada Sánchez** (Profesor de la USC, colaborador en el campo de la predicción medioambiental y en muestreo)

-**Cesar Sánchez Sellero** (Profesor de la USC, experto en temas relacionados con los tests de bondad de ajuste)

-**María José Lombardía** (Profesora de la USC, experta en poblaciones finitas y modelización con efectos aleatorios)

-**Alberto Rodríguez Casal** (Profesor de la USC, experto en estimación de conjuntos y en inferencia no paramétrica)

-**José María Matías Fernández** (Profesor de la UV, experto en técnicas de aprendizaje y en estadística computacional)

-**Ana Pérez González** (Profesora de la UV, experta en inferencia con datos faltantes)

-**Rubén Fernández Casal** (Profesor de la UDC, experto en estadística espacial y espacio temporal)

-**María Castellano Méndez** (Profesora de la USC, experta en diversos campos del consulting estadístico y con tesis doctoral en vías de presentación)

-**Rosa María Crujeiras Casais** (Contratada Post-doctoral de la USC para el consulting estadístico y experta en el campo de la estadística espacial y espacio temporal)

-**Beatriz Pateiro López** (Contratada de la USC para el consulting estadístico, con tesis doctoral muy avanzada en estimación de conjuntos, y muy conocedora del campo financiero).

-**Juan Carlos Reboredo** (Profesor de la USC, experto en fundamentos del análisis económico, especialmente en el contexto financiero)

-**Pedro Galeano San Miguel** (Investigador Parga Pondal de la USC experto en el análisis de observaciones influyentes)

-**Maria Piñeiro Lamas** (Contratada como técnico de investigación en el CIBER de Epidemiología y Salud Pública que actualmente realiza su tesis doctoral en el campo de la predicción medioambiental)
-**Isabel Fuentes** (Profesora de la UDC que actualmente realiza su tesis doctoral en el campo de los procesos puntuales)

Investigadores internacionales:

-**Winfried Stute** (Investigador de reconocido prestigio internacional en diversos campos. En particular, en tests de bondad de ajuste, procesos empíricos, matemática financiera, etc.)
-**Ingrid Van Keilegom** (Joven investigadora con un gran prestigio internacional logrado por sus recientes desarrollos en inferencia sobre curvas)
-**Stefan Sperlich** (Joven investigador con un gran prestigio internacional logrado por sus recientes desarrollos en inferencia sobre curvas)

Investigadores en formación:

-**Adela Martínez Calvo** (Becaria de FPI, que realiza su tesis en el campo de la estimación y predicción con datos funcionales)
-**Abelardo Monsalve Cobis** (Alumno del programa de doctorado interuniversitario de Estadística que realiza su tesis en el campo de la modelización con datos financieros)
-**Dos becarios de investigación** (Que podrían realizar su labor investigadora en diversos campos de investigación, como por ejemplo, Estadística Espacial, modelización de series financieras o en datos funcionales)
-**Un contratado a tiempo completo** Para realizar labores de ayuda en la gestión de todo el volumen de software que se puede desarrollar, de la administración informática de la web del grupo,....

y siguiendo las pautas de la sección anterior consideramos las siguientes actividades:

1) Modelización con efectos aleatorios.

Esta actividad estará coordinada por María José Lombardía y será realizada por los investigadores del grupo Wenceslao González Manteiga, María José Lombardía, José Manuel Prada Sánchez y César Sánchez Sellero.

Para las actividades detalladas en el punto 1.a) se colaborará con el catedrático Stefan Sperlich de la Universidad de Göttingen-Alemania (investigador internacional del grupo) y con la profesora María Dolores Martínez Miranda de la Universidad de Granada. La última actividad del punto 1.a) se desarrollará en colaboración con el Grupo Inditex.

Para las dos primeras actividades detalladas en el punto 1.b) se colaborará con el grupo de investigación dirigido por el catedrático Domingo Morales de la Universidad Miguel Hernández.

La última actividad del punto 1.b) se desarrollará en colaboración con el Instituto Galego de Estatística.

2) Inferencia en modelos de estadística espacial y espacio temporal.

Esta actividad estará coordinada por Manuel Febrero Bande. Los investigadores Rubén Fernández, Rosa Crujeiras y Wenceslao González, dada su experiencia previa, trabajarán en la extensión de las técnicas de estimación y contrastes en el dominio espectral a datos irregularmente espaciados. El desarrollo de métodos de contraste para modelos de estadística espacial y espacio temporal será desarrollado también por estos tres investigadores. A su vez, Rosa Crujeiras y Wenceslao González investigarán la posible utilización de los procesos empíricos de regresión para formular contrastes sobre la tendencia en un modelo de estadística espacial. En el campo de los procesos puntuales trabajarán Manuel Febrero e Isabel Fuentes. En todas las líneas, se realizará aplicación a datos reales (medioambientales, ecológicos y forestales).

3) Modelización con datos funcionales.

Actividad coordinada por Manuel Febrero Bande. Participarán en este campo Wenceslao González Manteiga, Manuel Febrero Bande, Adela Martínez Calvo y Pedro Galeano San Miguel. Con Adela se focalizará la investigación en la estimación y contrastes del modelo lineal funcional y con Pedro Galeano en el estudio de observaciones influyentes.

4) Estimación de conjuntos.

Beatriz Pateiro López y Alberto Rodríguez Casal (que actuará como coordinador de esta actividad) trabajarán a lo largo de los tres años del proyecto en los objetivos anteriores señalados en la sección anterior. Alberto Rodríguez Casal continuará con su colaboración con el profesor Antonio Cuevas. Beatriz Pateiro López trabajará en la finalización de su tesis doctoral en estimación de conjuntos.

5) Modelos no paramétricos y semiparamétricos .

Actuará como coordinador general de las diversas líneas de este campo el IP del proyecto. En los modelos de predicción medioambiental trabajarán los integrantes que actualmente desarrollan metodología para la Central Térmica de As Pontes : Wenceslao González Manteiga, José Manuel Prada Sánchez, Manuel Febrero Bande, Adela Martínez Calvo y María Piñeiro, que desarrolla su tesis doctoral en este contexto. Se incorpora al grupo de esta línea de investigación María Castellano por su experiencia previa en modelización con datos dependientes en problemas hidrología.

En tests de bondad de ajuste participarán Wenceslao González Manteiga, Cesar Sanchez Sellero y los investigadores internacionales : Winfried Stute e Ingrid Van Keilegom. Para la versión robustificada de los tests se trabaja actualmente con el grupo de Graciela Boente.

En el contexto de observaciones incompletas trabajarán Wenceslao González Manteiga y Ana Pérez González. Para datos sesgados se trabaja en colaboración con el grupo de Zaragoza liderado por José Antonio Cristóbal. Para datos censurados se trabaja con Ingrid Van Keilegom y estudiantes suyos.

En modelos parcialmente lineales con errores dependientes y sus aplicaciones económicas trabajarán Wenceslao González Manteiga y Juan Carlos Reboredo. Para el contexto de versiones robustas se trabaja con los investigadores Graciela Boente y Ana Bianco de la Universidad de Buenos Aires, expertos de rango internacional en Estadística Robusta.

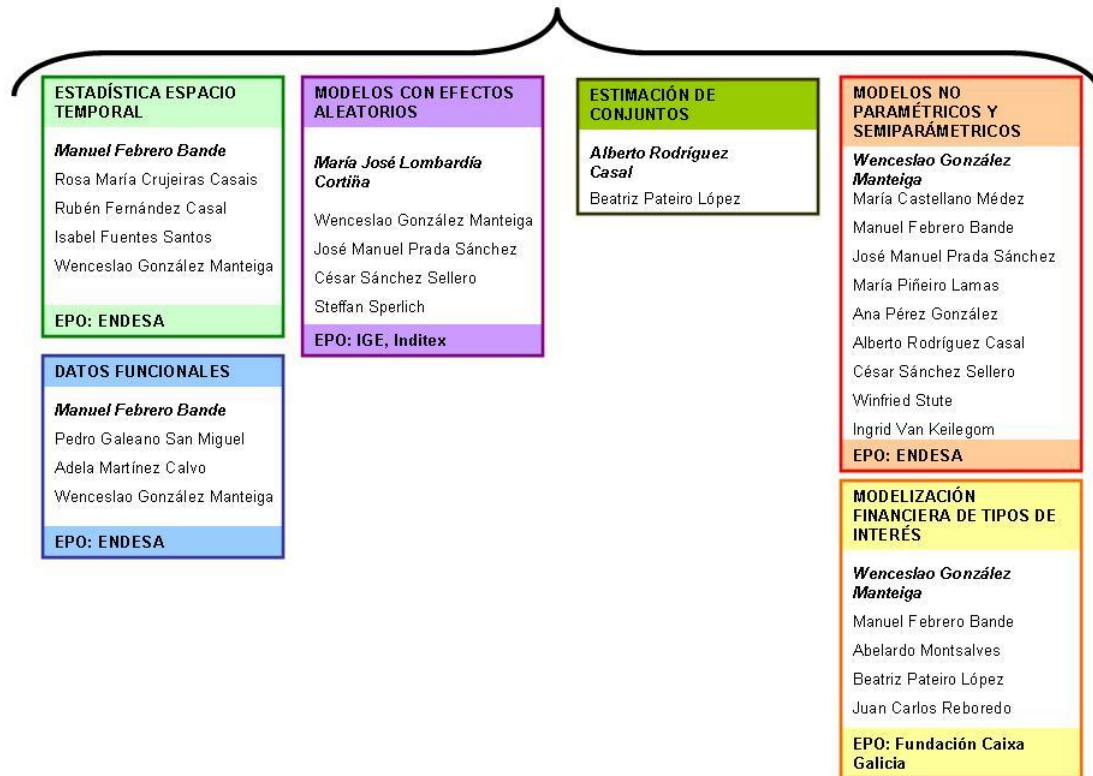
En SVMs trabajarán Wenceslao González Manteiga y José M. Matías. En este contexto jugará un papel importante el investigador Matías por las diversas aplicaciones en las que está involucrado.

6) Modelización financiera en series de tipos de interés.

En este campo participarán Wenceslao González Manteiga (que actuará como coordinador), Manuel Febrero Bande, Juan Carlos Reboredo y Beatriz Pateiro que ya disponen de una experiencia previa con sus estudios para la Fundación Caixa Galicia y se incorpora al grupo Abelardo Monsalve que desarrollará su tesis doctoral en este campo.

Se adjunta un gráfico orientativo de los subgrupos de trabajo y sus correspondientes coordinadores. A su vez las diversas líneas estarán conectadas estratégicamente con reuniones frecuentes de tipo general y complementarias a las que se desarrollarán con mucha más intensidad dentro de cada línea de trabajo.

WENCESLAO GONZÁLEZ MANTEIGA



4.1 MODELO DE CRONOGRAMA (ORIENTATIVO)

En este cronograma debe figurar la totalidad del personal investigador incluido en el formulario de solicitud y, en su caso, el personal contratado que se solicite con cargo al proyecto. Debe subrayarse el nombre de la persona responsable, en cada tarea.

Actividades/Tareas	Centro Ejecutor	Persona responsable y otras involucradas	Primer año (*)	Segundo año (*)	Tercer año (*)
Modelización con efectos aleatorios	Facultad de Matemáticas USC	María José Lombardía Cortiña W. González Manteiga J.M. Prada Sánchez C. Sánchez Sello S. Sperlich			
Inferencia en modelos de estadística espacial y espacio temporal	Facultad de Matemáticas USC Facultad de Informática UDC	Manuel Febrero Bande R. M. Crujeiras Casais R. Fernández Casal I. Fuentes Santos W. González Manteiga			
Modelización con datos funcionales	Facultad de Matemáticas USC	Manuel Febrero Bande P. Galeano Sanmiguel W. González Manteiga A. Martínez Calvo			
Estimación de conjuntos	Facultad de Matemáticas USC	Alberto Rodríguez Casal B. Pateiro López			

Modelos no paramétricos y semiparamétricos	Facultad de Matemáticas USC Facultad de Económicas UV	Wenceslao González Manteiga M. Castellano Méndez M. Febrero Bande A. Pérez González M. Piñeiro Lamas J.M. Prada Sánchez A. Rodríguez Casal C. Sánchez Sello W. Stute I. Van Keilegom			
Modelización financiera en series de tipos de interés	Facultad de Matemáticas USC Facultad de Económicas USC	Wenceslao González Manteiga M. Febrero Bande A. Monsalve Cobis B. Pateiro López J.C. Reboredo Nogueira			

(*) Colocar una X en el número de casillas (meses) que corresponda

5. BENEFICIOS DEL PROYECTO, DIFUSIÓN Y EXPLOTACIÓN, EN SU CASO, DE LOS RESULTADOS (máximo 1 página)

Deben destacarse, entre otros, los siguientes aspectos:

- Contribuciones científico-técnicas esperables del proyecto, beneficios esperables para el avance del conocimiento y de la tecnología y, en su caso, resultados esperables con posibilidad de transferencia ya sea a corto, medio o largo plazo.
 - Plan de difusión y, en su caso, de explotación, de los resultados del proyecto, que se valorará en el proceso de evaluación de la propuesta y en el de seguimiento del proyecto.
-

Para la difusión y explotación de resultados del proyecto, el grupo:

- 1) participará en conferencias y congresos de carácter nacional e internacional, con ponencias invitadas y organizando sesiones invitadas en congresos internacionales que en muchas ocasiones desembocan en publicaciones de números especiales de revistas de JCR.
- 2) con objeto de fortalecer y consolidar nuestra relación con el mundo empresarial, el grupo participará en foros de interacción con la industria, como ya se ha venido realizando a lo largo de los últimos años,
- 3) pondrá a disposición libre el software que pueda desarrollar, que será un elemento útil, tanto para las diversas aplicaciones con la consiguiente transferencia tecnológica, como para las investigaciones que se hagan en el futuro.
- 4) Tratará de mantener el alto volumen de publicaciones de gran calidad de los últimos años.

Puntualmente, por la importancia que podría tener en transferencia tecnológica, podríamos destacar especialmente las siguientes líneas:

A) Modelización con efectos aleatorios :

El desarrollo de nueva metodología en estimación con modelos mixtos, tiene especial interés en el campo de la medicina, en estudios sociales y económicos. En particular, el estudio de parámetros de interés (medias, totales, proporciones, etc.) en áreas pequeñas tiene interés para la Comunidad Autónoma de Galicia, que con mayor frecuencia demanda estudios a nivel comarcal. Posibles beneficiarios de estas investigaciones son claramente el Instituto Galego de Estatística y el Grupo Inditex con el que se inició recientemente una conexión investigadora.

B) Inferencia en modelos estadísticos de tipo espacial y espacio temporal:

Estudio de datos medioambientales o de tipo ecológico, predicción de niveles de SO₂, predicción de niveles de concentración de metales pesados en musgo, contaminación atmosférica,...etc. (Posibles beneficiarios: ENDESA y el Departamento de Biología Celular y Ecología de la Universidad de Santiago de Compostela).

C) Modelización con datos funcionales:

Estudio, estimación y predicción de niveles de concentración de contaminantes, relación de variables objetivables medibles en la cerámica con variables de mercado, etc. (Posible beneficiario: ENDESA, Valora Consultores de Gestión y ASCER).

D) Modelos no paramétricos y semiparamétricos:

De las diversas líneas de investigación en modelos no paramétricos y semiparamétricos destacamos el impacto que puede tener la modelización de la predicción medioambiental para ENDESA y los estudios con SVMs que son susceptibles de diversas aplicaciones industriales.

E) Modelización financiera en series de tipo de interés :

El claro beneficiario para este contexto es el entorno bancario. Especialmente la Fundación Caixa Galicia con la que se viene colaborando desde hace años.

6. HISTORIAL DEL EQUIPO SOLICITANTE EN EL TEMA PROPUESTO

En el apartado 2 de este informe de solicitud se explica con detalle la relación entre el trabajo previo de los miembros del equipo y el proyecto de investigación que se presenta aquí. Para resumir y concretar de manera objetiva el trabajo realizado por los miembros del equipo solicitante, se puede señalar que en el período 2001-2007, los integrantes de este grupo han publicado un total de más de 100 artículos de investigación en revistas internacionales incluidas en los Journal Citation Reports. Todos ellos relacionados con alguna de las líneas de investigación a desarrollar en el proyecto. Omitimos aquí por restricciones de espacio, artículos publicados en revistas que no están en el JCR, capítulos de libros o proceedings y las numerosas aportaciones en distintos congresos de rango nacional o internacional, muchas de ellas por invitación, y nos limitamos a presentar una selección de las publicaciones muy relevantes en los distintos campos de las líneas del proyecto. En el impreso de solicitud tipo C se adjunta la lista completa de artículos en JCR.

Trabajos de investigación de ámbito internacional publicados por los miembros del equipo en el período 2001-2007.

- 1) Akritas, M.G. y **Van Keilegom, I.** (2001), "Nonparametric ANCOVA methods for heterocedastic nonparametric regression models". JASA, 96, 220-232.
- 2) Akritas, M.G. y **Van Keilegom, I.** (2001), "Nonparametric estimation of the residual distribution". Scandinavian Journal of Statistics, 28, 549-568.
- 3) Akritas, M.G. y **Van Keilegom, I.** (2003), "Estimation of the bivariate and marginal distributions with censored data". J.Royal Statist. Soc.-Series B, 65, 457-471.
- 4) Aneiros Pérez, G., **González Manteiga, W.** y Vieu, P. (2004), "Estimation and testing in a partial linear regression model under long-memory dependence". Bernoulli, 10, 1, 49-78
- 5) Cadarso Suárez, C., Roca Pardiñas, J., Figueiras, A. y **González Manteiga, W.** (2004), "Nonparametric estimation of the odds ratios for continuous exposures using generalized additive models with an unknown link function". Statistics in Medicine, 24, 1169-1184.
- 6) Cao, R. y **González Manteiga, W.** (2006), "Empirical likelihood tests for two-sample problems via nonparametric density estimation". Canadian Journal of Statistics. 34, 61-77.
- 7) Cao, R. y **Van Keilegom, I.** (2007), "Goodness of fit tests for conditional models under censoring and truncation". Journal of Econometrics. Por aparecer.
- 8) **Castellano Méndez, M., González Manteiga, W., Febrero Bande, M., Prada-Sánchez, J.M.** y Lozano-Calderón, R. (2004), "Modelling of the monthly and daily behaviour of the runoff of the Xallas river using Box-Jenkins and neural networks methods". Journal of Hydrology, 296, 38-58.
- 9) Chacón-Durán, J. E. y **Rodríguez-Casal, A.** (2005) "On the L1 consistency of wavelet density estimates". Canadian Journal of Statistics. 33, 489-496.
- 10) Chen, X., Linton, O. y **Van Keilegom, I.** (2003), "Estimation of semiparametric models when the criterion function is not smooth". Econometrica, 71, 1591-1608.
- 11) Claeskens, G. y **Van Keilegom, I.** (2003), "Bootstrap confidence bands for regression curves and their derivatives. Ann. of Statistics. 31, 1852-1884.
- 12) Crujeiras, R.M., Fernández-Casal, R., González-Manteiga, W. (2007) "Comparing spatial dependence structures using spectral density estimators". Environmetrics. 18, 793-808.
- 13) Cuevas, A. y **Rodríguez Casal, A.** (2004) "On boundary estimation". Advances in Applied Probability. 36, 340-354.
- 14) Cuevas, A., **Febrero Bande, M.** y Fraiman, R. (2001). "Cluster analysis: a further approach based on density estimation". Computational Stat. & Data Analysis. 36, p.441-459.
- 15) Cuevas, A., **Febrero Bande, M.** y Fraiman, R. (2002). "Linear functional regression: the case of fixed design functional response". The Canadian Journal of Statistics. 30,2, 285-300.
- 16) Cuevas, A., **Febrero, M.** y Fraiman, R. (2004), "An anova test for functional data". Computational Statistics & Data Analysis. 47, 111-122.
- 17) Cuevas, A., **Febrero, M.** y Fraiman, R. (2006) "On the use of bootstrap for estimating functions with functional data". Computational Statistics and Data Analysis, 51, nº 2, 1063-1074.
- 18) Cuevas, A., Fraiman, R. y **Rodríguez Casal, A.** (2007) "A nonparametric approach to the estimation of lengths and surface areas". Annals of Statistics. 35, 1031-1051.
- 19) Cuevas, A., **González-Manteiga, W.** y **Rodríguez-Casal, A.** (2006) "Plug-in estimation of general level sets". Australian and New Zealand Journal of Statistics. 48, 7-19.
- 20) De Uña-Álvarez, J. y **Rodríguez-Casal, A.** (2007) "Nonparametric estimation from length-biased data under competing risks". Computacional Statistics and Data Analysis. 51, 2653-2669.

- 21) Delgado, M. y **González Manteiga, W.** (2001) "Significance testing in nonparametric regression based on the bootstrap". *Ann. of Stat.* 29,5, 1469-1507.
- 22) Dette, H., Neumeyer, N. y **Van Keilegom, I.** (2007) "A new test for the parametric form of the variance function in nonparametric regression". *J. Royal Stat. Soc. Series B*, 69, 903-917.
- 23) Du, Y., Akritas, M.G. y **Van Keilegom, I.** (2003), "Nonparametric methods for analysis of covariance with censored data". *Biometrika*, 90, 269-287.
- 24) Einmahl, J. y **Van Keilegom, I.** (2007), "Specification tests in nonparametric regression". *Journal of Econometrics*. Por aparecer.
- 25) Einmahl, J. y **Van Keilegom, I.** (2007), "Tests for independence in nonparametric regression". *Stat. Sinica*. Por aparecer.
- 26) El Ghouch, A. y **Van Keilegom, I.** (2007), "Nonparametric regression with dependent censored data". *Scand. Journal of Stat.* Por aparecer.
- 27) El Ghouch, A., **Van Keilegom, I.** y McKeague, I. (2007), "Empirical likelihood confidence intervals for dependent duration data". *Econometric Theory*. Por aparecer.
- 28) **Febrero, M., Galeano, P. y González Manteiga, W.** (2007) "Outlier detection in functional data by depth measures with application to identify abnormal NOx levels". *Environmetrics*, Por aparecer.
- 29) **Fernández Casal, R., González Manteiga, W. y Febrero Bande, M.** (2003), "Flexible spatio-temporal stationary variogram models". *Statistics and Computing*, 13, 127-136.
- 30) Fernández de Castro, B. y **González Manteiga, W.** (2007), "Boosting for real and functional samples. An application to an environmental problem". *Stoch. Envir. Res. And Risk. Ass.* por aparecer.
- 31) Fernández de Castro, B., Guillas, S. y **González Manteiga, W.** (2005), "Functional samples and bootstrap for predicting SO2 levels". *Technometrics*, 47, 2, 212-222.
- 32) **Galeano, P.** (2007). "The use of cumulative sums for detection of changepoints in the rate parameter of a Poisson Process". *Computational Statistics and Data Analysis* 51, 6151-6165.
- 33) **Galeano, P.;** Peña, D. and Tsay, R. S. (2006). "Outlier Detection in Multivariate Time Series by Projection Pursuit". *Journal of the American Statistical Association* 101, 654-669.
- 34) **González Manteiga, W. y Pérez González, A.** (2004), "Goodness of fit test for linear regression models with missing response data". *The Canadian Journal of Statistics*, 34, 1, 1-22.
- 35) **González Manteiga, W.**, Martínez Miranda, M.D. y **Pérez González, A.** (2004). "The choice of smoothing parameter in nonparametric regression through Wild Bootstrap". *Computational Statistics & Data Analysis* 47, 487-515.
- 36) **González-Manteiga, W., Lombardía, M.J.**, Molina, I., Morales, D. y Santamaría, L. (2007). Estimation of the Mean Squared Error of Predictors of Small Area Parameters under a Logistic Mixed Model. *Computational Statistics & Data Analysis*. 51, 2720-2733.
- 37) Hall, P. y **Van Keilegom, I.** (2003), "Using difference-based methods for inference in nonparametric regression with time series errors". *J. Royal. Statist. Soc. -Series B*, 65, 443-456.
- 38) Hall, P. y **Van Keilegom, I.** (2004), "Testing for monotone increasing hazard rate". *Ann. Statist.* 33, 1109-1137.
- 39) Hall, P. y **Van Keilegom, I.** (2007), "Two-sample tests in functional data analysis starting from discrete data". *Stat Sinica*. 17, 1511-1532.
- 40) Härdle, W., Huet, S., Mammen, E., **Sperlich, S.** (2004) "Bootstrap Inference in Semiparametric Generalized Additive Models". *Econometric Theory*, 20, 265-300.
- 41) Härdle, W., **Sperlich, S.** y Spokoiny, V. (2001) "Structural tests in additive regression". *J. Am. Statist. Assoc.*, 96, 1333-1347.
- 42) Hengartner, N.W., **Sperlich, S.** (2005) "Rate Optimal Estimation with the Integration Method in the Presence of Many Covariates". *Journal of Multivariate Analysis*, 95(2), 246-27
- 43) Heuchenne, C. y **Van Keilegom, I.** (2007), "Location estimation in nonparametric regression with censored data". *J. Multiv. Anal.*, 98, 1558-1582.
- 44) Heuchenne, C. y **Van Keilegom, I.** (2007), "Nonlinear regression with censored data". *Technometrics*, 49, 34-44.
- 45) Hjort, N.L., McKeague, I.W. y **Van Keilegom, I.** (2007), "Extending the scope of empirical likelihood". *Ann. Of Stat.* Por aparecer.
- 46) Li, G. y **Van Keilegom, I.** (2002), "Likelihood ratio confidence bands in nonparametric regression with censored data". *Scandinavian Journal of Statistics*, 29, 547-562.
- 47) Linton, O., **Sperlich, S. y Van Keilegom, I.** (2007), "Estimation of a semiparametric transformation model". *Ann of Stat.* Por aparecer.
- 48) **Lombardía, M.J., González-Manteiga, W. y Prada-Sánchez, J.M.** (2005). "Estimation of a finite population distribution function based on a linear model with unknown heteroscedastic errors". *Canadian Journal of Statistics*. 33, 181-200.
- 49) **Lombardía, M.J., Sperlich, S.** (2007). "Semiparametric Inference in Generalized Mixed Effects Models". *Journal of the Royal Statistical Society, Series B*. Por aparecer
- 50) Martínez Miranda, M.D., Raya Miranda, R., **González Manteiga, W.** y González Carmona, A.. (2007). "A bootstrap local bandwidth selector for additive models". *Journal of Computational and Graphical Statistics*. Por aparecer.
- 51) **Matías, J.M.,** Vaamonde, A., Taboada, J. y **González Manteiga, W.** (2004), "Comparison of Kriging and Neural Networks with application to the exploitation of a slate mine" *Mathematical Geology*, 36, 4, 463-486.
- 52) **Matías, J.M.,** Vaamonde, A., Taboada, J. y **González Manteiga, W.** (2004), "Support vector machines and gradient boosting for graphical estimation of a slate deposit". *Stochastic Environment al Research and Risk Assessment*. 18, 309-323.

- 53) **Matías, J.M., y González Manteiga, W.** (2006), "Regularized kriging as a generalization of simple, universal and Bayesian Kriging". *Stochastic Environmental Research and Risk Assessment*. 20, 243-258.
- 54) Menezes, R.; García-Soidán, P. y **Febrero, M.** (2007) "A kernel variogram estimator for clustered data". *Scandinavian Journal of Statistics* doi:10.1111/j.1467-9469.2007.00566.x.
- 55) Molina, I., Saei, A and **Lombardía, M.J.** (2007). "Estimation of Unemployment Rates in Small Areas by a Multinomial Logit Mixed Model". *Journal of The Royal Statistical Society Series A*. 170, 975-1000.
- 56) Neumeyer, N., **Sperlich, S.** (2006) Comparison of Separable Components in Different Samples. *Scandinavian Journal of Statistics*, 33, 477 - 501.
- 57) Nielsen, J.P., **Sperlich, S.** (2005) Smooth backfitting in practice. *Journal of the Royal Statistical Society, B*, 67(1), 43-61 Note: elected for the list of highlights in 2005, entered the list of most cited papers during the last three years (top 10)
- 58) **Pardo Fernández, J.C. y Van Keilegom, I.** (2006), "Comparison of regression curves with censored responses". *Scand. Journal of Stat.* 33, 409-434.
- 59) **Pardo Fernández, J.C., Van Keilegom, I. y González Manteiga, W.** (2007), "Testing for the equality of k regression curves". *Statistica Sinica*. 17, 3, 1115-1137.
- 60) **Pardo Fernández, J.C., Van Keilegom, I. y González Manteiga, W.** (2007), "Goodness-of-fit tests for parametric models in censored regression". *Canadian Journal of Statistics* 35, 2, 249-264.
- 61) Roca Pardiñas, Cadarso Suárez, C. y **González Manteiga, W.** (2005), "Testing for interactions in generalized additive models: Application to SO2 pollution data". *Statistics and computing*. 15, 289-299.
- 62) Roca Pardiñas, J., **González Manteiga, W., Febrero Bande, M., Prada Sánchez, J.M.** y Cadarso Suárez, C. (2004), "Predicting binary time series of SO2 using generalized additive models with unknown link function". *Environmetrics*. 15, 729-742.
- 63) Roca Pardiñas, J., **Sperlich, S.** (2007) "Testing the link when the index is semiparametric – A comparison study". *Comp. Stat. & Data Analysis*, 12, 6565 - 6581.
- 64) **Rodríguez-Casal, A.** (2007) "Set estimation under convexity type assumptions". *Annales de l'I.H.P. - Probabilités & Statistiques*. 43, 763-774.
- 65) Rodríguez-Póo, J., **Sperlich, S.**, Fernández, A.I. (2005) "Semiparametric Three Step Estimation Methods for Simultaneous Equation Systems". *Journal of Applied Econometrics*, 20, 699-721.
- 66) Rodríguez-Póo, J., **Sperlich, S.**, Vieu, P. (2003) "Semiparametric Estimation of Weak and Strong Separable Models". *Econometric Theory*, 19, 1008-1039.
- 67) **Sánchez Sellero, C., González Manteiga, W. y Van Keilegom, I.** (2004), "Uniform representation of product-limit integrals with applications". *Scandinavian Journal of Statistics*. 32, 563-581.
- 68) **Stute, W.** (2001), "Residual analysis for ARCH(p)-Time series." *TEST*, 10, 393-403.
- 69) **Stute, W.** y Ferreira, E. (2004), "Testing for differences between conditional means in a time series context". *JASA*, 99, 169-174.
- 70) **Stute, W.** y Zhu, L.X. (2002), "Model checks for generalized linear models" *Scandinavian Journal of Statistics*. 29, 535-545.
- 71) **Stute, W.** y Zhu, L.X. (2005), "Nonparametric checks for single-index models". *Annals of Statistics*. 33, 1048-1083.
- 72) **Stute, W.**, Koul, H. y Li, F. (2005), "Model diagnosis for SETAR time series". *Statistica Sinica*. 15, 795-817.
- 73) **Stute, W.**, Xu, W.L. y Zhu, L.X. (2007), "Model diagnosis for parametric regression in high dimensional spaces". *Biometrika*, por aparecer.
- 74) **Stute, W.**, Xue, L. y Zhu, L.X. (2007), "Empirical likelihood inference in nonlinear error-in-covariables models with validation data". *JASA*. 102, 332-346.
- 75) **Stute, W.** y Delgado, M. (2007), "Specification tests of conditional distributions". *Journal of Econometrics*. Por aparecer.
- 76) **Van Keilegom I.**, Akritas, M.G. y Veraverbeke, N. (2001), "Estimation of the conditional distribution in regression with censored data". *Computational Statistics & Data Analysis* 35, 487-500.
- 77) **Van Keilegom, I. González Manteiga, W. y Sánchez Sellero, C.** (2007), "Goodness of fit tests in parametric regression based on the estimation of the error distribution". *TEST*, por aparecer.
- 78) **Van Keilegom, I.** y Carroll, R.J. (2007), "Backfitting versus profiling in general criterion functions". *Statistica Sinica*. 17, 797-816.
- 79) **Van Keilegom, I.** y Hettmansperger, T.P. (2002), "Inference on multivariate M-estimators based on bivariate censored data". *JASA*, 97, 328-336.
- 80) **Van Keilegom, I.** y Veraverbeke, N. (2002), "Density and hazard estimation in censored regression models". *Bernoulli*, 8, 607-625.
- 81) Vilar Fernández, J.M. Vilar Fernández, J.A. y **González Manteiga, W.** (2001), "Bootstrap tests for nonparametric comparison of regression curves with dependent errors". *TEST*. 16,1, 123-144.
- 82) Wang, L. y **Van Keilegom, I.** (2007), "Nonparametric test for the form of parametric regression with time series errors". *Stat. Sinica*, 17, 369-386.

6.1. FINANCIACIÓN PÚBLICA Y PRIVADA (PROYECTOS Y CONTRATOS DE I+D) DE LOS MIEMBROS DEL EQUIPO INVESTIGADOR

Debe indicarse únicamente lo financiado en los últimos cinco años (2003-2007), ya sea de ámbito autonómico, nacional o internacional.

Deben incluirse también las solicitudes pendientes de resolución.

Título del proyecto o contrato	Relación con la solicitud que ahora se presenta (1)	Investigador Principal	Subvención concedida o solicitada	Empresa/administración financiadora y referencia del proyecto	Periodo de vigencia o fecha de la solicitud (2)
			EUROS		
Mejora del SIPEI-2003	0	Wenceslao González Manteiga	28.850	ENDESA Generación	C
Ampliación del análisis estadístico de variables estáticas y dinámicas relacionadas con el funcionamiento eléctrico en los sistemas de coladas de hornos de ferroaleaciones	2	Wenceslao González Manteiga	18.000	Ferroatlántica I+D	C
Análise e predición de activos financeiros	1	Wenceslao González Manteiga	33.657	Fundación Caixa Galicia – Claudio San Martín	C
Estadística de gasto del turismo que pernocta en establecimientos hoteleros 2003	2	José Manuel Prada Sánchez	75.991,38	Instituto Galego de Estatística	C
Asesoramiento en el diseño y obtención de datos en una encuesta a la población del municipio de Arnoia (Ourense) sobre las actitudes ciudadanas respecto al desarrollo sostenible en el municipio	2	César Andrés Sánchez Sello	450	IDOM S.A.	C
Estudio estadístico sobre el turismo estival en Galicia. Encuesta de destino 2003	2	José Manuel Prada Sánchez	26.128,09	Turgalicia S.A.	C
Estudio estadístico sobre el excursionismo en Galicia. Excursionismo 2003	2	José Manuel Prada Sánchez	8.259,65	Turgalicia S.A.	C
Estudio estadístico sobre el turismo en Galicia. Encuesta de origen 2003	2	José Manuel Prada Sánchez	8.307,42	Turgalicia S.A.	C

SIPE2004	0	Wenceslao González Manteiga	29.000	ENDESA Generación	C
Técnicas estadísticas orientadas al control de variables estáticas y dinámicas relacionadas con el funcionamiento eléctrico en hornos de ferroaleaciones	1	Wenceslao González Manteiga	18.000	Ferroatlántica I+D	C
Análise e predición de activos financeiros (continuación)	1	Wenceslao González Manteiga	29.450	Fundación Caixa Galicia – Claudio San Martín	C
Determinación del tamaño muestral en análisis de caracterización de residuos de envases ligeros	2	César Andrés Sánchez Sello	1.200	Novotec Consultores S.A.	C
Estudio estadístico sobre el turismo estival en Galicia. Encuesta de destino 2004	2	José Manuel Prada Sánchez	26.990,32	Turgalicia S.A.	C
Estudio estadístico sobre el excursionismo en Galicia. Excursionismo 2004	2	José Manuel Prada Sánchez	9.065,55	Turgalicia S.A.	C
Estudio estadístico sobre el turismo en Galicia. Encuesta de origen 2004	2	José Manuel Prada Sánchez	9.511,56	Turgalicia S.A.	C
SIPEI 2005	0	Wenceslao González Manteiga	29.000	ENDESA Generación	C
Análise e predición da estrutura temporal dos tipos de xuro	0	Wenceslao González Manteiga	25.543	Fundación Caixa Galicia – Claudio San Martín	C
Ampliación de técnicas estadísticas orientadas al control de variables estáticas y dinámicas relacionadas con el funcionamiento eléctrico y los sistemas de coladas en hornos de ferroaleaciones	2	Wenceslao González Manteiga	18.500	Ferroatlántica I+D	C

Estadística de gasto del turismo que pernocta en establecimientos hoteleros en Galicia durante 2005	2	José Manuel Prada Sánchez	80.462,10	Instituto Galego de Estatística	C
Estudio estadístico de caracterización de residuos farmacéuticos del sistema SIGRE	2	César Andrés Sánchez Sello	500	Novotec Consultores S.A.	C
Diseño de la encuesta y análisis de datos sobre las actitudes ciudadanas respecto del desarrollo sostenible dentro de la Agenda 21 de la comarca de Pontevedra (Ayuntamientos de Barro, Pontecalderas y Vilaboa)	2	César Andrés Sánchez Sello	600	IDOM S.A.	C
Asesoramiento estadístico para el control de calidad de materiales recuperados de envases ligeros	2	César Andrés Sánchez Sello	1200	Novotec Consultores S.A.	C
Desarrollo estadístico-redes neuronales para la CTCC Besós	1	Wenceslao González Manteiga	6.600	ENDESA Servicios	C
Estudio estadístico sobre el turismo estival en Galicia. Encuesta de destino 2005	2	José Manuel Prada Sánchez	27.773,03	Turgalicia S.A.	C
Estudio estadístico sobre el excursionismo en Galicia. Excursionismo 2005	2	José Manuel Prada Sánchez	9.328,46	Turgalicia S.A.	C
Estudio estadístico sobre el turismo en Galicia. Encuesta de origen 2005	2	José Manuel Prada Sánchez	9.844,46	Turgalicia S.A.	C
SIPEI 2006	0	Wenceslao González Manteiga	29.000	ENDESA Generación	C
Programas de formación, de integración multicultural de los inmigrantes y de colaboración en materia de estadística sobre migraciones	2	Wenceslao González Manteiga	8.500	Secretaría Xeral de Emigración (Xunta de Galicia)	C

Convenio de colaboración entre el Instituto Galego de Estadística y la Universidad de Santiago de Compostela para la realización de un proyecto de investigación en muestreo en áreas pequeñas, con aplicaciones	0	Wenceslao González Manteiga y María José Lombardía Cortiña	36.000	Instituto Galego de Estatística	C
SIPEI 2007	0	Wenceslao González Manteiga	30.000	ENDESA Generación	C
Asesoramiento estadístico para el diseño de una metodología de determinación del contenido de envases en los residuos urbanos	2	César Andrés Sánchez Sello	3.300	APPLUS, Norcontrol S.L.U.	C
Estudio estadístico sobre o volume de turismo e excursionismo en Galicia durante 2006. Enquisas de destino, orixe e excursionismo.	2	José Manuel Prada Sánchez	53.987,84	Turgalicia S.A.	C
Realización do deseño metodolóxico e análise de resultados de enquisas que permitan cuantificar o volume de turismo e excursionismo en Galicia en 2007	2	José Manuel Prada Sánchez	55.067,60	Turgalicia S.A.	C
Deseño dun modelo que explique o impacto no gasto sanitario da dispersión xeográfica e o envellecemento da poboación	1	Manuel Febrero Bande	6.034,48	Servizo Galego de Saúde	C
SIPEI 2008	0	Wenceslao González Manteiga	36.000	ENDESA Generación	C
Análise estatística de problemas socioeconómicos, medioambientais e tecnolóxicos	0	Alberto Rodríguez Casal	52.900	Dirección Xeral de Investigación, Desenvolvemento e Innovación	C
Plataforma estadística GRID de computación paralela y distribuida	1	Manuel Febrero Bande	64.000	Consellería de Innovación e Industria. Xunta de Galicia	C
Inferencia Estadística Funcional. Nuevas Aplicaciones a las Finanzas, al Medioambiente, a la Industria y a la Economía	0	Wenceslao González Manteiga	163.030	Ministerio de Educación y Ciencia. Fondos FEDER	C

Statistical Analysis os Association and Dependence in Complex Data. Interuniversity Attraction Pole (IAP)	0	Ingrid Van Keilegom	25.000	Belgiam Science Policy	C
Grupos de referencia competitiva. Modelización estatística e aplicacións	1	Wenceslao González Manteiga	340.000	Consellería de Educación e Ordenación Universitaria	C
Red MATHEMATICA Consulting and Computing de Galicia	1	Wenceslao González Manteiga	180.000	Consellería de Educación e Ordenación Universitaria	C
PGIDIT07PXIB207031PR	0	Wenceslao González Manteiga	31.800	Consellería de Innovación e Industria	C

- (1) Escribase 0, 1, 2 ó 3 según la siguiente clave: 0 = es el mismo tema; 1 = está muy relacionado; 2 = está algo relacionado; 3 = sin relación
- (2) Escribase una C o una S según se trate de una concesión o de una solicitud.

7. CAPACIDAD FORMATIVA DEL PROYECTO Y DEL EQUIPO SOLICITANTE

(en caso de Proyecto Coordinado deberá rellenarse para cada uno de los equipos participantes que solicite becas)

Este apartado sólo debe rellenarse si se ha respondido afirmativamente a la pregunta correspondiente en el cuestionario de solicitud.

Debe justificarse que el equipo solicitante está en condiciones de recibir becarios (del Subprograma de Formación de Investigadores) asociados a este proyecto y debe argumentarse la capacidad formativa del equipo.

Nota: el personal necesario para la ejecución del proyecto deberá incluirlo en el apartado de personal del presupuesto solicitado. La concesión de becarios FPI, sólo será posible para un número limitado de los proyectos aprobados, en función de la valoración del proyecto y de la capacidad formativa del equipo.

La capacidad formativa del equipo viene avalada por el gran número de doctores que se han formado y siguen formándose con nosotros a lo largo de los últimos años, algunos de ellos incorporados en este grupo de investigación como doctores. Además, desde nuestro grupo de investigación siempre hemos considerado importante implicar a jóvenes investigadores en formación. En esta ocasión, Adela Martínez Calvo realiza su tesis doctoral bajo la dirección del IP en el campo de datos funcionales y Abelardo Monsalve (estudiante venezolano becado por el gobierno de Venezuela), que ha comenzado a trabajar en modelos financieros dentro de nuestro programa de doctorado con Mención de Calidad del MEC (que en este momento es Programa Oficial de Postgrado). Además, varios miembros del equipo investigador tienen experiencia en la dirección de tesis doctorales (en particular, el IP del proyecto ha dirigido 22 tesis) y no cabe duda de que cualquiera de las líneas de investigación contempladas en este proyecto es un campo de proyección futura, donde un becario podría formarse, no sólo metodológicamente, si no también enfrentándose a la aplicación práctica de los métodos estudiados en problemas reales de índole industrial, financiero, medioambiental,... Por otra parte, las consolidadas conexiones internacionales del grupo con expertos de los distintos campos que se abordan en el proyecto facilitarían (como ya ha ocurrido con anteriores becarios) la movilidad de nuestros becarios y el incremento de su nivel formativo.