**MODELOS DE INFERENCIA NO PARAMÉTRICA PARA INDICADORES MEDIOAMBIENTALES**

1. **Documento 1B-1: Resumo do proxecto**

A estatística semiparamétrica multidimensional, a análise de datos direccionais e a modelización con datos funcionais constitúen campos de máximo interese no entorno da Estatística na investigación actual a nivel internacional. No presente proxecto de investigación preténdese realizar novos desenrolos metodolóxicos en cada un destes campos e con especial énfases na aplicación ó contexto do Medio Ambiente, contribuíndo a determinar estratexias para preservar o entorno. Moitos dos novos métodos que se desenrolen serán implementados por primeira vez ou representarán xeneralizacións de versións anteriores, e serán validados mediante resultados teóricos e de simulación. Todas as técnicas propostas presentan un grande atractivo tanto desde o punto de vista académico, xa que a investigación no eido dos datos funcionais e direccionais son áreas de investigación emerxente, coma desde o punto de vista empírico, posto que na actualidade en diferentes campos das ciencias aplicadas (enxeñería medioambiental, química, biométrica, medicina,...) se recolectan grandes cantidades de datos o que fai necesario a formulación multidimensional dos problemas ou a consideración de que os datos son curvas. Os métodos obtidos ó longo deste proxecto poderán ser aplicados no contexto do Medio Ambiente grazas a ampla colaboración entre o Departamento de Estatística e Investigación Operativa da Universidade de Santiago de Compostela (USC) e a Central Térmica de As Pontes, propiedade de Endesa Generación S.A.

Objetivos:

Prediccion niveles SO2 y validación:

Clasificación fuente inmisión:

Desarrollo de software que integra el estado del arte actual en datos funcionales

1. **Documento 1B-2: Estado da arte da investigación proposta**

Fai unhas décadas era impensable que o desenrolo económico puidera afectar tan negativamente á natureza como para chegar a representar un serio problema. É por isto que nos últimos anos os gobernos e as organizacións ecoloxistas en conxunción cos científicos buscan solucións para prever a degradación medioambiental. En particular as accións políticas para a protección do medioambiente obrigan ás empresas a desenrolar plans de control mediambiental de calidade de aire, auga e solo.

A Estatística xoga un papel moi importante á hora de predicir os niveis de polución que as industrias causan no seu entorno. Existen diversas técnicas, de tipo paramétrico e non paramétrico, para a predición de valores futuros. Dentro da clase dos modelos paramétricos para a predición da tendencia inclúense os clásicos ARMA, que engloban como casos particulares ós puramente autorregresivos (AR) e ós modelos de medias móbiles (MA) (Wei, 1990;Brockwell e Davis, 1987). Estes modelos foron xeneralizados ó caso multidimensional dando lugar ós modelos VARMA (Lütkelpohl, 2005) e, posteriormente, ós VECM (Engle e Granger, 1987; Lütkelpohl, 2005) que teñen en conta a posible relación de cointegración entre as series. Por outra banda, tamén se desenvolveron modelos non paramétricos para a estimación da tendencia como os modelos aditivos (Hastie e Tibshirani, 1990), modelos baseados en redes neuronais (Tsay, 2002) ou a estimación tipo núcleo (Hastie e Tibshirani, 1990).

Na predición de indicadores medioambientais, así como no análise de datos procedentes de diversas ciencias, como a Bioloxía, a Xeoloxía ou a Meteoroloxía, tamén pode ser interesante ter en conta datos de tipo direccional, como por exemplo a dirección do vento. A maioría dos avances teóricos na estatística de variables aleatorias direccionais é relativamente recente. A referencia chave na teoría desta área é o libro de Mardia (2000) así como o de Fisher (1993). Ademais estanse a publicar numerosos traballos como o de Zhao e Chengqing (2001) que utilizan o Teorema Central do Límite para o erro cuadrático integrado por medio da estimación tipo núcleo da densidade esférica, ampliando así o traballo de Hall (1984).

O progreso das ferramentas de computación, tanto en termos de memoria como de capacidade, permite traballar con grandes conxuntos de datos. En particular, é habitual atoparnos cunha variable aleatoria observada en continuo, polo que un pode chegar a considerar que o que observa é realmente a realización dunha variable funcional. Isto fixo que nos últimos anos os datos funcionais estean de plena actualidade, e revistas de máximo nivel como Statistica Sinica (Vol.14,Nº3,2004) ou Computational Statistics & Data Analysis (Vol.51,10,2007) publicaran números monográficos adicados a desenvolvementos metodolóxicos con datos funcionais e aplicacións. O libro que máis contribuíu a popularizar as técnicas estatísticas para datos funcionais é sen dúbida algunha o publicado por Ramsay e Silverman (2005) que considera o caso dos modelos de regresión con datos funcionais a través de procedementos próximos ó suavizado spline. O outro hito bibliográfico é o de Ferraty e Vieu (2006) que trata os modelos de regresión mediante a extensión da estimación tipo núcleo. Febrero (2010) trata as medidas de influencia do modelo funcional lineal con resposta escalar no que a previsión é funcional. Anteriormente xa se propuxeron varios enfoques para a estimación deste modelo, como por exemplo Cardot, Ferraty e Sarda (2003) ou Ramsay e Silverman (2005) que analizaron o uso de funcións de forma restrinxida e métodos de penalización. Recentemente deseñáronse modelos máis flexibles que permiten conxugar a estimación da compoñente funcional con outras covariables que non o son (Aneiros e Vieu, 2009; Baíllo e Grané, 2009). El avance logrado en los últimos años correspondide a la creación y optimización de nuevo software científico. Indudablemente el paquete fda (Ramsay et al. 2010) es la referencia básica para trabarar en R. Los autores del grupo STAPH mantienen activa la web <http://lsp.ups-tlse.fr/staph> con el código R de sus funciones.

A lexislación de calidade de aire obriga ás industrias a controlar os niveis de contaminación que causan no seu entorno como se comentou ó inicio. En particular, na Central Térmica de As Pontes implantouse un sistema suplementario da contaminación atmosférica. Dito sistema sufriu diversas evolucións ó longo dos anos con diferentes modelos de predición de tipo semiparamétrico do SO2 dos que cabe destacar:

1. Un modelo de predición non paramétrico con erros ARI, García Jurado e outros (Technometrics, 1995)
2. Modelización de tipo semiparamétrico espacio-temporal, Angulo e outros (Environmental and Ecological Statistics, 1998)
3. Modelización con modelos parcialmente lineais, Prada Sánchez e outros (Environmetrics, 2000)
4. Modelización usando redes neuronais, Fernández de Castro e outros (Air & Waste Manage, 2003).
5. Predición con modelos de resposta binaria de tipo semiparamétrico, Roca Pardiñas e outros (Environmetrics, 2004).
6. Modelización con modelos funcionais, Fernández de Castro e outros (Technometrics, 2005)
7. Estimación e contrastes con modelos aditivos xeneralizados, Roca Pardiñas e outros (Statistics and Computing,2005)
8. Predición utilizando técnicas de aprendizaxe tipo boosting, Fernández de Castro e González Manteiga (Stoch. Environ. Res. Risk. Asses., 2007)

Na actualidade vívese un momento apaixonante dende o punto de vista científico. A Central Térmica sufriu un proceso de adaptación que permite o seu funcionamento con 100% carbón de importación, as emisións de SO2 reducíronse en torno a un 95% e empezou a funcionar unha nova central de ciclo combinado onde a serie de NOx cobra vital importancia. Estamos ante o desafío de predicir unha serie con resposta bidimensional onde os erros bidimensionais poden ter distinta estrutura paramétrica e teñamos por tanto que xogar con aspectos de cointegración. Na central xa está implantado un sistema que predí, con media hora de antelación, os niveis de SO2 e de NOx de forma independente. Estase traballando na modelización semiparamétrica con estrutura de correlación nos erros e os primeiros resultados xa foron presentados en congresos nacionais (SEIO 2009) e internacionais (ISI 2009). Ademais, deseñouse a primeira versión dunha ferramenta que permite decidir cal é a orixe dun episodio de alteración da calidade do aire: o ciclo combinado, a central térmica ou outros posibles focos de contaminación, na que xoga un papel moi importante unha covariable de tipo direccional: a dirección do vento.

O noso grupo de investigación tamén abordou outros problemas no eido medioambiental ademais do da central térmica. Por exemplo, Febrero e outros (2008) estudaron os niveis anómalos de NOx na cidade de Barcelona mediante medidas de profundidade funcional.

1. **Documento 1B-3: Obxectivos do proxecto**

Neste proxecto existen tres vía de investigación principais. A continuación detállanse os obxectivos concretos a acadar en cada unha delas:

1. Elaboración de meteoroloxía dentro do campo dos datos funcionais.
   1. Estudio do modelo de regresión con covariables funcionais e resposta real o funcional.
      1. Modelo lineal xeneralizado (GLM)
      2. Modelo parcialmente lineal (LMP)
      3. Estudio aditivo xeneralizado (GAM)
   2. Extensión aos modelos con:
      1. Datos dependentes.
      2. Datos espacio-temporais.
      3. Datos direccionais.
      4. Datos faltantes
   3. Estudio de simulación e validación dos modelos anteriores.
   4. Aplicación dos modelos 1.1.1., 1.1.2. y 1.1.3. á predición de series temporais e espacio-temporais medioambientais con datos direccionais.
2. Elaboración de metodoloxía dentro do campo dos modelos de series temporais
   1. Modelización semiparamétrica multidimensional con estrutura de correlación nos erros.
   2. Estimación, contrastes e predición en ditos modelos.
   3. Aplicación ó sistema de predición na Central Térmica das Pontes para as variables SO2 e NOx.
3. Elaboración de meteoroloxía dentro do campo de datos direccionais.
   1. Estudio de distribucións notables de datos direccionais.
   2. Estudio de modelos de predición con datos direccionais.
   3. Aplicación do estudado en 3.1. e 3.2. á predición de series medioambientais.
4. Elaboración de software
   1. Transferencia tecnolóxica da na metodoloxía á Central Térmica de As Pontes.
   2. Creación e optimización do novo software científico creado a partir do estudio desenrolado no apartado.

Anéxase un diagrama indicativo das posibles relacións entre as distintas liña de traballo.

1. **Documento 1B-4: Aplicabilidade e posible transferencia dos resultados**

Da frutífera relación entre a Sección de Medio Ambiente da U.P.T. As Pontes e o Departamento de Estatística e Investigación Operativa da Universidade de Santiago de Compostela nace o Sistema de Predición Estatística de Inmisión (SIPEI) debido al compromiso de Endesa co medio ambiente e o seu interese por mellorar este sistema de predición.

O Sistema de Predición de Estatística de Inmisión utiliza datos de inmisión minutales dispoñibles no sistema MEDAS para obter predicións dos valores de dióxido de xofre y de óxidos de nitróxeno, con media hora de antelación, usando modelos aditivos. Este sistema de predición tamén proporciónanos una predición probabilística de superación de los niveles de SO2 y NOx; tendo en conta os límites contemplados na lexislación vixente.

Ademais, SIPEI calcula a probabilidade estimada de se é a Central Térmica ou o Ciclo Combinado o causante do episodio de alteración da calidade do aire.

Nos últimos anos, o Sistema de Predición Estatística de inmisión adoptou unha estrutura modular na que se poden distinguir tres grandes bloques: a conexión, a predición e a visualización.

A conexión engloba todo o que ten que ver co orixe, a estrutura e a manera de obter os datos; constitue a unión entre o provedor dos datos e o propio sistema de predición. Neste módulo obtense os datos orixinais, formatéanse convenientemente y póñense a disposición del módulo de predición.

A predición fai referencia aos aspectos máis metodolóxicos, matemáticos y estatísticos; aquí encóntranse os procesos xurdidos da implementación dos modelos desenrolados. Neste módulo lense os datos formateados, aplícanse os modelos, realízanse os cálculos, obtense os resultados que se poñen a disposición do módulo de visualización.

A visualización comporta aspectos relacionados coa disposición e presentación da información. Neste módulo lense os resultados, xenáranse os gráficos e os informes e ponse toda a información a disposición do usuario final.

De todo isto conclúese co sistema é unha importante ferramenta de información e axuda na toma de decisións.

Anéxase unha imaxe coa pantalla principal da visualización.

1. **Documento 1B-5: Plan de difusión e explotación de resultados**

Os avances logrados durante a execución deste proxecto estarán ao alcance da comunidade investigadora e da sociedade en xeral, a través da publicación de Technical Reports (accesibles desde a páxina web do departamento http://eio.usc.es/pub/reports.html), de comunicacións en congresos, proxecto Final de Master en Estatística e das tese doutorais dos bolseiros integrantes do equipo.

Tamén se publicarán artigos en revistas internacionais incluídas nos Journal Citation Reports tanto nas de carácter máis metodolóxico (Scandinavian Journal of Statistics, Canadian Journal of Statistics, Statistica Sinica…) como nas máis aplicadas (Technometrics, Environmetrics, Econometrica, Journal of Chemometrics, Computational Statistics and Data Analysis…). Por outra banda, dous membros do equipo pertencen a ERCIM (European Research Consortium for Informatics and Mathematics, http://www.ercim.org/) como coordinadores, xunto con Frederic Ferraty e Philippe Vieu da Universite Paul Sabatier de Toulouse, do tópico Statistics for Functional Data dentro do grupo de traballo Computing & Statistics (http://www.dcs.bbk.ac.uk/ercim/TrackSFD.html). Ademais o departamento pertence á rede europea IAP (http://www.stat.ucl.ac.be/IAP/PhaseVI/index2.html) que organiza workshops, seminarios, cursos,... Todo isto favorecerá a difusión internacional dos resultados obtidos e do software creado.

Para a construción do software empregarase a linguaxe de programación R (software libre) , ferramenta empleada para a creación dun paquete de datos funcionais que estará a disposisión de futuros usuarios no repositorio de contribuciones (http://cran.proyect.org). Ademáis aproveitando que o departamento pertence ao programa CONSOLIDER, incluirase no repositorio de software que se ten previsto crear dentro dito programa.

Todos os resultados que se van obter permitirán o estudo de datos medioambientais ou de tipo ecolóxico, a predición da evolución de indicadores medioambientais,... o que favorecerá a transferencia tecnolóxica de todos estes avances teóricos ao mundo empresarial (Posibles beneficiarios: ENDESA).

1. **Documento 1B-6: Interese para o sector e posibles beneficios**

Tanto a temática dos datos funcionais como a temática dos datos direccionais son de recente aparición polo que hai un amplo abanico de liñas de investigación aínda sen explorar.

Por iso o principal obxectivos deste proxecto será estender os modelos de predición xa construídos no ámbito unidimensional e multidimensional ao terreo funcional, é de grande interese desde o punto de vista teórico. Outro objetivo es clasificar as fontes de inmisión dos niveles de SO2 y NOx para coñecer con maior precisión e rapidez si el orixe foi da Central, do ciclo combinado ou doutra fonte exóxeno a Central.

Ademais esta metodoloxía pode proporcionar ferramentas para o tratamento de grandes cantidades de datos de carácter funcional de diversas disciplinas (Economía, Medio Ambiente, Medicina, Química,...), que os modelos existentes non son capaces de tratar de maneira adecuada.

Por outra banda, este proxecto tamén ten como obxectivo levar os datos direccionais o campo das series de tempo, o que é de gran interese tanto dende un punto de vista teórico como práctico. Ademais, isto sería beneficioso á hora de construír modelos de predición xa que podería incluír variables meteorolóxicas (por exemplo, dirección e velocidade do vento) ó longo do tempo para perfeccionar sistemas de predición ambientais.

1. **Documento 1B-7: Metodoloxía**

No que ao desenvolvemento matemático se refire, a metodoloxía a seguir será a habitual nestes casos empregando tanto ferramentas estatísticas como probabilísticas: deseño dun modelo con datos direccionais en series de tempo e estudo do seu comportamento con datos reais de control de calidade do aire, construción dun novo estimador para a regresión funcional e estudo do seu comportamento mostral e da súa distribución asintótica; elaboración de tests baseados en dito estimador e validación do seu comportamento mediante simulación; extensión das técnicas de aprendizaxe aos datos funcionais…

En canto ao aspecto computacional, o software que se constrúa estará escrito na linguaxe de programación R (software libre). O entorno de programación estatística R consolidouse como unha plataforma de desenrolo de algoritmos estatísticos avanzados. O entorno é libre e gratuíto o que o fai accesible á comunidade científica, académica e profesional os resultados das aplicacións e métodos implementados. As técnicas e tratamentos básicos xa están implementados no repositorio de contribuciones (<http://cran.proyect.org>). Ademais, ao traballar con grandes conxuntos de datos, sobre os que se farán continuas remostraxes, xerarase unha carga computacional considerable que fará imprescindible o emprego de equipos informáticos veloces e con gran capacidade.

No relativo á metodoloxía de traballo propiamente dita, os sete membros do grupo reuniranse semanalmente. Por outra banda, as viaxes que unha das bolseiras realiza dúas veces á semana a Central Térmica de As Pontes, permitirán a transferencia tecnolóxica dos traballos realizados á área medioambiental, e a estreita relación entre o Departamento de Estatística e IO da USC.

Cómpre salientar que a relación entre o grupo e a Sección de Medio Ambiente da Central Térmica de As Pontes (a través de varios proxectos entre ENDESA e a USC, como se pode observar nos currículums dos membros do equipo) xa propiciou a transferencia tecnolóxica ao longo dos últimos anos (anéxase unha carta do xefe da Sección de Medio Ambiente da Central Térmica). Mostra disto é a aplicación de numerosas metodoloxías estatísticas para a predición de niveis de SO2 como se amosa no diagrama que se anexa.

1. **Documento 1B-8: Plan de traballo**

O traballo do proxecto está planificado para unha duración de tres anos. A continuación detállanse as actividades a realizar en cada un deles.

Primeiro ano.

Levarase a cabo unha profunda revisión das metodoloxías relativas ás series temporais e, datos funcionais e datos direccionais. Para isto consultarase a bibliografía citada no apartado do Estado da Arte, así como os traballos que se vaian publicando ao longo da execución do proxecto. Isto corresponde ao cumprimento dos obxectivos 1.1 a 1.2, 2.1 a 2.3 e 3.1.

Segundo ano.

Deseñarase unha nova clase de estimadores para a regresión funcional a partir de estimadores de mínima distancia con respecto a un estimador piloto nonparamétrico de tipo núcleo funcional. Elaboraranse, en base a estas distancias, tests para a selección de modelos de regresión e realizaranse simulacións para a súa validación. Ademais, estenderanse diversos modelos para series temporais e espacio-temporais ao campo funcional. Isto corresponde ao cumprimento dos obxectivos 1.3 a 1.6 e 2.4. Tamén se aplicaran os modelos de predición con datos direccionais, obxetivo 3.2.

Os algoritmos implementados se colgaran de la web del proyecto . Ademaís o paquete de datos funcionais **desarrollado** se colgará no repositorio de contribuciones (<http://cran.proyect.org>).

Terceiro ano.

Aplicaranse os resultados obtidos durante os dous anos anteriores a datos medioambientais. Ademais o software construído favorecerá a transferencia tecnolóxica a empresas como ENDESA. Isto corresponde ao cumprimento dos obxectivos 1.7, 2.5 e 3.3.

Todos os membros do equipo participarán activamente nas tarefas anteriormente mencionadas.

Por outra banda, ao longo de todo o proxecto teranse reunión periódicas con representantes de ENDESA , o que mellorará a aplicabilidade dos traballos realizados.

Anéxase un cronograma coas tarefas previstas para os tres anos de duración do proxecto.

1. **Documento 1B-11: Explicación e xustificación dos gastos**

A continuación xustifícanse os gastos que aparecen na Memoria Económica.

MATERIAL INVENTARIABLE OU BIBLIOGRÁFICO.

1. Actualización de 7 ordenadores persoais dos membros do equipo e do equipo informático instalado na Central Térmica de As Pontes. Adquisición dunha matriz de discos duros.

Como se expón no apartado de Outros Medios da Memoria Científico-Técnica as necesidades computacionais do proxecto esixen renovar os equipos informáticos dos cinco membros do grupo e do equipo informático instalado na Central Térmica de As Pontes, e adquirir un disco duro común de varios terabytes que permita almacenar as grandes cantidades de datos coas que se vai traballar.

2. Bibliografía

Como se especifica no Plan de Traballo, durante o primeiro ano do proxecto revisarase minuciosamente as metodoloxías relativas ás series temporais e aos datos funcionais existentes. Para isto é imprescindible que o grupo adquira libros de máximo interese nestes campos (como algúns dos citados no apartado do Estado da Arte) así como todos os que se vaian publicando durante os próximos anos.

3. Tradución e edición de artigos.

Este gasto xurde do feito de que moitas das revistas nas que se publicarán os resultados obtidos ao longo deste proxecto esixen que os autores aboen os gastos de edición dos seus traballos.

MATERIAL FUNXIBLE

1. Consumibles, gastos en material de impresora e papel.

Neste apartado inclúese todos os gastos dos membros do equipo relativos a material de oficina.

DESPRAZAMENTOS

1. Desprazamentos a congresos, foros e seminarios (unha viaxe ao ano por membro).

Considérase de gran importancia a participación dos membros do grupo en congresos, foros e seminarios que favorezan a difusión dos resultados obtidos durante o proxecto así como o acceso ás metodoloxías que outros grupos de investigación estean a desenvolver en campos afíns tanto aos datos funcionais e ás series de tempo, como ás súas aplicacións medioambientais e financeiras.

2. Dúas viaxes semanais á Central Térmica de As Pontes.

Unha das bolseiras do grupo viaxa dúas veces á semana á Central Térmica de As Pontes, o que permite dirixir as aplicacións das metodoloxías estudadas cara a resolución de problemas e necesidades reais do mundo empresarial, favorecendo a transferencia tecnolóxica dos resultado obtidos.

3. Visitas trimestrais de todo o grupo á Central Térmica de As Pontes.

Ademais das viaxes semanais especificadas no apartado anterior, os cinco membros do grupos visitarán varias veces ao ano á Central Térmica de As Pontes.

1. **Documento 1B-12: Experiencia do equipo de investigación**

Tanto o IP Wenceslao González coma José M. Prada e Manuel Febrero, teñen ampla experiencia en diversas liñas de investigación: Inferencia en modelos espazo temporais (aplicación na predición medioambiental con mapas de SO2 e dirección de vento, predición xeolóxica, modelos ecolóxicos, etc.), Modelización econométrico-financeira, Técnicas de aprendizaxe (aplicación no control industrial, aplicacións medioambientais na predición de niveis de SO2, etc.), tests de bondade de axuste para modelos diversos, Inferencia con datos funcionais (aplicación á predición de series de tipo de interese, de curvas de ozono, etc.), … Ademais, tamén son integrantes de varias redes de investigación tanto a nivel nacional (Nodo CESGA, programa CONSOLIDER) como internacional (IAP Network), e do proxecto de investigación de convocatoria pública: “Inferencia Estadística funcional. Nuevas aplicaciones a las Finanzas, al Medioambiente, a la Industria y a la Economía” (2005-2008); IP: Wenceslao González Manteiga; subvencionado polo Ministerio de Educación y Ciencia. Por razóns de espazo omítense a participación en proxectos con empresas, capítulos de libros ou proceedings, e numerosas aportacións en distintos congresos de rango nacional e internacional. Ademáis María Piñeiro, Adela Martinez e Manuel Oviedo teñen experiencia en la participación en outros proxectos de investigación a nivel autonómico, nacional e internacional como se pode consultar os currículums

Só se citan algúns dos artigos máis importantes e recentes (para máis información consultar os currículums).

1)Cao, R.; González-Manteiga, W. (To appear). "Goodness-of-fit tests for conditional models under censoring and truncation". Journal of Econometrics.

2)Febrero-Bande, M.; Galeano, P. e González-Manteiga, W. (To appear). "A Functional Analysis of NOx levels: location and scale estimation and outlier detection". Computational Statistics Aceptado para publicación.

3)Cadarso Suárez, Carmen María e González-Manteiga, W. (2007 (in press)). "Statistics in Biomedical Research". ARBOR.

4)Roca Pardiñas, Javier; Cadarso Suárez, Carmen María e González-Manteiga, W. (2007). "Bootstrap methods for testing interactions in GAMs. Applications". In: Statistical Methods for Biostatistics and Related Fields (Härdle W, Mori Y, and Vieu Ph. Eds.), 147-166. Springer.

5)Menezes, R.; García-Soidán, P.H. e Febrero-Bande, M. (2007). "A kernel variogram estimator for clustered data". Scandinavian Journal of Statistics In press.

6)Cuevas, A.; Febrero-Bande, M. e Fraiman, R. (2007). "Robust estimation and classification for functional data via projection-based depth notions". Computational Statistics In press.

7)Cuevas, A.; González-Manteiga, W. e Rodríguez-Casal, A. (2006). "Plug-in estimation of general level sets.". Australian and New Zealand Journal of Statistics. 48, 7-19. Blackwell.

8)González-Manteiga, W.; Lombardía Cortiña, María José; Molina Perala, I.; Morales González, D. e Santamaría, L. (2006). "Estimation of the Mean Squared Error of Predictors of Small Area Parameters under a Logistic Mixed Model". Computational Statistics & Data Analysis Journal In Press. Elsevier Science Bv.

9)Pateiro-López, B. e González-Manteiga, W. (2006). "Multivariate Partially Linear Models". Statistics and Probability Letters 76, Issue 14, 1543-1549. Elsevier.

10)Cuevas, A.; Febrero-Bande, M. e Fraiman, R. (2006). "On the use of the bootstrap for estimating functions with functional data". Computational Statistics & Data Analysis 51, nº 2, 1063-1074.

11)Cadarso Suárez, Carmen María; Roca Pardiñas, Javier; Figueiras Guzmán, Adolfo e González-Manteiga, W. (2005). "Nonparametric estimation of the Odds Ratios for Continuous Exposures Using Generalized Additive Models with an Unknown Link Function". Statistics in Medicine vol. 24, nº 8, pp. 1169-1184. Wiley.

12)Fernández de Castro, Belén M.; Guillas, S. e González-Manteiga, W. (2005). "Functional Samples and Bootstrap for Predicting SO2 Levels.". Technometrics. vol. 47, n. 2., p. 212-222..

13)Roca Pardiñas, Javier; Cadarso Suárez, Carmen María e González-Manteiga, W. (2005). "Testing for interactions in Generalized Additive Models:application to SO2 pollution". Statistics and Computing, 15,, 289-299,.

14)Menezes, R.; García-Soidán, P.H. e Febrero-Bande, M. (2005). "A comparison of approaches for valid variogram achievement". Computational Statistics 20 4, 623-642. Springer.

15)Lombardía Cortiña, María José; González-Manteiga, W. e 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. Statistical Society of Canada.

16)García-Soidán, P.H.; Febrero-Bande, M. e González-Manteiga, W. (2004). "Nonparametric kernel estimation of an isotropic variogram.". Journal of Stat. Planning Inference. 121, 65-92. Elsevier.

17)Cuevas, A.; Febrero-Bande, M. e Fraiman, R. (2004). "An anova test for functional data". Computational Statistics and Data Analysis 47, 111-122. Elsevier.

18)Vilar Fernández, Juan Manuel e González-Manteiga, W. (2004). "Nonparametric comparison of curves with dependent errors". Statistics.

19)Aneiros Pérez, German; González-Manteiga, W. e Vieu, Philippe (2004). "Estimation and testing in a partial linear regression model under long memory dependence". Bernoulli.

20)Matias Fernández, José María; Vaamonde Liste, Antonio; Taboada Castro, Javier e González-Manteiga, W. (2004). "Comparison of kriging and neural networks with application to the exploitation of a slate mine". Mathematical Geology.

21)González-Manteiga, W.; Martinez Miranda, Maria Dolores e Pérez González, Ana (2004). "The choice of smoothing parameter in nonparametric regression through wild bootstrap". Computational Statistics & Data Analysis.

22)Roca Pardiñas, Javier; González-Manteiga, W.; Febrero-Bande, M.; Prada-Sánchez, J. M. e Cadarso Suárez, C. M. (2004). "Predicting binary time series of SO2 using generalized additive models with unknown link function". Environmetrics 15, p. 729-742.

1. **Documento 1B-13: Instalacións e instrumentos**

O departamento de Estatística e Investigación Operativa da USC, do que forman parte os membros do equipo, posúe un grid de computación paralela obtido na recente convocatoria de infraestructuras da Xunta de Galicia que consta de 11 servidores AMD opteron 2.4 GHz con dobre procesador dobre núcleo. Dito grid pode satisfacer as necesidades de simulación froito da execución deste proxecto.

Ademais cada un dos membros dispón dun ordenador persoal (cunha antigüidade media de dous anos).

Por outra banda, a bolseira do grupo que visita semanalmente a Central Térmica de As Pontes emprega alí un equipo informático pertencente ao departamento co que xestiona os datos medioambientais proporcionados pola Central (cunha antigüidade de 5 anos).

1. **Documento 1B-14: outros medios**

Debido as necesidades de velocidade e alta capacidade computacionais que precisa o correcto desenvolvemento do proxecto, téñense que renovar os cinco equipos informáticos dos membros grupo.

Ademais, sería preciso dispoñer dunha matriz de discos duros común para almacenar as inxentes cantidades de datos que se van a manexar e optimizar a utilización do grid do que xa dispón o departamento.

Polos mesmos motivos expostos anteriormente, tamén se precisa renovar o equipo informático que se ten nas instalacións da Central Térmica de As Pontes.