

ANEXO VI

 XUNTA DE GALICIA CONSELLERÍA DE INNOVACIÓN E INDUSTRIA Dirección Xeral de Investigación, Desenvolvemento e Innovación		 Innovación, Ciencia e Tecnoloxía
---	---	---

PROGRAMA DE PROMOCIÓN XERAL DA INVESTIGACIÓN

Ano 2007

MODELO PXI-02

PROCEDEMENTO AXUDAS PARA PROXECTOS DE INVESTIGACIÓN BÁSICA	CÓDIGO DO PROCEDEMENTO IN825 B	DOCUMENTO SOLICITUDE
--	--	--------------------------------

DATOS DO/A SOLICITANTE (INVESTIGADOR/A PRINCIPAL DO EQUIPO)

PRIMEIRO APELIDO GONZALEZ	SEGUNDO APELIDO MANTEIGA	NOME WENCESLAO	NIF 32618861P
SEXO HOME <input checked="" type="checkbox"/> MULLER <input type="checkbox"/>	ORGANISMO UNIVERSIDADE DE SANTIAGO DE COMPOSTELA		DATA DE LECTURA DA TESE 27/11/1982
CENTRO FAC. DE MATEMATICAS			CÓDIGO* 207
DEPARTAMENTO ESTADÍSTICA E INVESTIGACION OPERATIVA			CÓDIGO* S36
ENDEREZO RUA LOPE GÓMEZ DE MARZO A S/N CAMPUS SUR		LOCALIDADE SANTIAGO	CONCELLO SANTIAGO DE COMPOSTELA
CÓDIGO POSTAL 15782	PROVINCIA A CORUNA	TELÉFONO 981563100	EXTENSIÓN 13204
FAX 981597054		CORREO ELECTRÓNICO wenceslao@usc.es	

*VER LISTAXE DE CÓDIGOS DE CENTROS E DEPARTAMENTOS

DATOS DO PROXECTO

TÍTULO MODELOS DE PREDICION CON DATOS FUNCIONAIS. APLICACIONS MEDIOAMBIENTAIS E FINANCEIRAS.				
TIPO DE AXUDA				
<input checked="" type="checkbox"/> MODALIDADE A: PROXECTOS DE INVESTIGACIÓN <input type="checkbox"/> MODALIDADE B: PROXECTOS COORDINADOS				
NO CASO DE PROXECTOS COORDINADOS INDICAR:				
	1º APELIDO	2º APELIDO	NOME	ORGANISMO
COORDINADOR/A XERAL DO PROXECTO (INVESTIGADOR/A PRINCIPAL SUBPROXECTO 1):				
INVESTIGADOR/A PRINCIPAL SUBPROXECTO 2:				
INVESTIGADOR/A PRINCIPAL SUBPROXECTO 3:				
INVESTIGADOR/A PRINCIPAL SUBPROXECTO 4:				
OBSERVACIÓNS:				

DURACIÓN <input type="checkbox"/> 1 ANO <input type="checkbox"/> 2 ANOS <input checked="" type="checkbox"/> 3 ANOS	ÁREA TEMÁTICA DO PROXECTO MTM	CÓDIGO DA ANEP 265	CÓDIGO DA ÁREA DE CONECEMENTO 265
---	----------------------------------	-----------------------	--------------------------------------

RESUMO DE CUSTOS

	ANUALIDADE 2007	ANUALIDADE 2008	ANUALIDADE 2009	ANUALIDADE 2010	SUMA
CONTRATACIÓN DE PERSOAL	0	0	0	0	0
MATERIAL INVENTARIABLE E BIBLIOGRÁFICO	5.375	6.375	6.375	6.375	24.500
MATERIAL FUNXIBLE	1.000	1.000	1.000	1.000	4.000
AXUDAS DE CUSTO POR DESPRAZAMENTO	7.675	7.675	7.675	7.675	30.700
OUTROS GASTOS	0	0	0	0	0
TOTAL	14.050	15.050	15.050	15.050	59.200
IMPORTE DO PROXECTO (1)	59.200	CUSTOS INDIRECTOS (15% de (1)) (2)		8.880	
		SUBVENCIÓN SOLICITADA (1)+(2)		68.080	

Contía máxima: 120.000 € para un período máximo de 3 anos.

VISTO E PRACE DA AUTORIDADE QUE REPRESENTA LEGALMENTE AO ORGANISMO/CENTRO PÚBLICO

Dña. / D. SENÉN BARRO AMENEIRO	SIGNATURA E SELO
CARGO RECTOR	
ORGANISMO UNIVERSIDADE DE SANTIAGO DE COMPOSTELA	
C.I.F. Q 1518001	

LEXISLACIÓN APLICABLE Orde da Consellería de Innovación e Industria do 20 de abril de 2007 pola que se convocan axudas do Programa de Promoción Xeral de Investigación do Plan Galego de Investigación, Desenvolvemento e Innovación Tecnolóxica (INCITE) para o ano 2007.	(Para cubrir pola Administración) RECIBIDO REVISADO E CONFORME	NÚMERO DE EXPEDIENTE DATA DE ENTRADA / / DATA DE EFECTOS / / DATA DE SAÍDA / /
SINATURA DO/A SOLICITANTE SANTIAGO DE COMPOSTELA, 19 de xuño de 2007		

Conselleiro de Innovación e Industria

Os impresos válidos para a presentación obtense no enderezo <http://www.xunta.es/cumprimentacion-electronica>

AIB 1 IN825B

cod_web: 388

32618861P

32618861P

cod_web: 388

AIB 1 IN825B



AXUDAS PARA A REALIZACIÓN DE PROXECTOS DE INVESTIGACIÓN

DOCUMENTO Nº 1-A

32618861P

cod_web: 388

IN825B
AIB 1

INVESTIGADOR PRINCIPAL

Documento 1A-1

Investigador Principal (doutor que deberá pertencer a algunha das categorías que figura no apartado 2.3.a) do Anexo I, que deberá ter dedicación a tempo completo ao correspondente organismo. No caso de centros sanitarios, pode ter dedicación a tempo parcial.

SINATURAS DE CONFORMIDADE(*)

PRIMEIRO APELIDO GONZALEZ	SEGUNDO APELIDO MANTEIGA	NOME WENCESLAO	N.I.F. 32618861P	ORGANISMO UNIVERSIDADE DE SANTIAGO DE COMPOSTELA	SEXO <input checked="" type="checkbox"/> HOME <input type="checkbox"/> MULLER	CV
VINCULACIÓN AO ORGANISMO NUMERARIO	POSTO DESEMPEÑADO CATEDRÁTICO DE UNIVERSIDADE	DEDICACIÓN AO ORGANISMO TEMPO COMPLETO	TITULACIÓN DOUTOR EN MATEMATICAS	DATA LECTURA DA TESE 27/11/1982	DEDICACIÓN AO PROXECTO UNICA	

OUTROS MEMBROS DO EQUIPO, CONFORME AO ESTABLECIDO NO PUNTO 2.3, APARTADOS a), b), c) E d) DA ORDE DE CONVOCATORIA:

PRIMEIRO APELIDO FEBRERO	SEGUNDO APELIDO BANDE	NOME MANUEL	N.I.F. 34961128R	ORGANISMO UNIVERSIDADE DE SANTIAGO DE COMPOSTELA	SEXO <input checked="" type="checkbox"/> HOME <input type="checkbox"/> MULLER	CV
VINCULACIÓN AO ORGANISMO NUMERARIO	POSTO DESEMPEÑADO PROFESOR TITULAR DE UNIVERSIDADE	DEDICACIÓN AO ORGANISMO TEMPO COMPLETO	TITULACIÓN DOUTOR EN MATEMATICAS			
DATA LECTURA DA TESE 15/09/1995	DEDICACIÓN AO PROXECTO UNICA	TIPO DE PERSOAL PERSOAL INVESTIGADOR ESTABLE				

PRIMEIRO APELIDO PRADA	SEGUNDO APELIDO SANCHEZ	NOME JOSE MANUEL	N.I.F. 34602477N	ORGANISMO UNIVERSIDADE DE SANTIAGO DE COMPOSTELA	SEXO <input checked="" type="checkbox"/> HOME <input type="checkbox"/> MULLER	CV
VINCULACIÓN AO ORGANISMO NUMERARIO	POSTO DESEMPEÑADO CATEDRÁTICO DE UNIVERSIDADE	DEDICACIÓN AO ORGANISMO TEMPO COMPLETO	TITULACIÓN DOUTOR EN MATEMATICAS			
DATA LECTURA DA TESE	DEDICACIÓN AO PROXECTO UNICA	TIPO DE PERSOAL PERSOAL INVESTIGADOR ESTABLE				

O INVESTIGADOR PRINCIPAL DO EQUIPO DECLARA QUE SON CERTOS OS DATOS ARRIBA INDICADOS

SINATURA DO INVESTIGADOR/A PRINCIPAL

(*)Coa súa sinatura, o membro do equipo dá a súa conformidade para a participación no proxecto e promete que os datos contados no seu currículo son certos.

OUTROS MEMBROS DO EQUIPO, CONFORME AO ESTABLECIDO NO PUNTO 2.3, APARTADOS a), b), c) , d) E e) DO ANEXO I:

SINATURAS DE CONFORMIDADE(*)

PRIMEIRO APELIDO MARTINEZ	SEGUNDO APELIDO CALVO	NOME ADELA	N.I.F. 44825032X	ORGANISMO UNIVERSIDADE DE SANTIAGO DE COMPOSTELA	SEXO <input type="checkbox"/> HOME <input checked="" type="checkbox"/> MULLER	CV
VINCULACIÓN AO ORGANISMO (1) CONTRATO/BOLSEIRO PREDOUTORAL OU EQUIVALENTE		POSTO DESEMPEÑADO BOLSEIRA PREDOUTORAL	DEDICACIÓN AO ORGANISMO TEMPO COMPLETO		TITULACIÓN LICENCIADO EN MATEMATICAS	
DATA LECTURA DA TESE		DEDICACIÓN AO PROXECTO UNICA		TIPO DE PERSOAL PERSOAL INVESTIGADOR EN FORMACION		

PRIMEIRO APELIDO PINEIRO	SEGUNDO APELIDO LAMAS	NOME MARIA	N.I.F. 44828699C	ORGANISMO UNIVERSIDADE DE SANTIAGO DE COMPOSTELA	SEXO <input type="checkbox"/> HOME <input checked="" type="checkbox"/> MULLER	CV
VINCULACIÓN AO ORGANISMO (1) CONTRATO/BOLSEIRO PREDOUTORAL OU EQUIVALENTE		POSTO DESEMPEÑADO BOLSEIRA DE CONTRATOS E CONVENIOS	DEDICACIÓN AO ORGANISMO TEMPO COMPLETO		TITULACIÓN LICENCIADO EN MATEMATICAS	
DATA LECTURA DA TESE		DEDICACIÓN AO PROXECTO UNICA		TIPO DE PERSOAL PERSOAL INVESTIGADOR EN FORMACION		

PRIMEIRO APELIDO	SEGUNDO APELIDO	NOME	N.I.F.	ORGANISMO	SEXO <input type="checkbox"/> HOME <input type="checkbox"/> MULLER	
VINCULACIÓN AO ORGANISMO (1)		POSTO DESEMPEÑADO	DEDICACIÓN AO ORGANISMO		TITULACIÓN	
DATA LECTURA DA TESE		DEDICACIÓN AO PROXECTO		TIPO DE PERSOAL		

PRIMEIRO APELIDO	SEGUNDO APELIDO	NOME	N.I.F.	ORGANISMO	SEXO <input type="checkbox"/> HOME <input type="checkbox"/> MULLER	
VINCULACIÓN AO ORGANISMO (1)		POSTO DESEMPEÑADO	DEDICACIÓN AO ORGANISMO		TITULACIÓN	
DATA LECTURA DA TESE		DEDICACIÓN AO PROXECTO		TIPO DE PERSOAL		

(1) No caso de presentarse como investigador asociado, indicar a vinculación ao seu respectivo organismo.

O INVESTIGADOR PRINCIPAL DO EQUIPO DECLARA QUE SON CERTOS OS DATOS ARRIBA INDICADOS

SINATURA DO INVESTIGADOR/A PRINCIPAL

(*)Coa súa sinatura, o membro do equipo dá a súa conformidade para a participación no proxecto e promete que os datos contados no seu currículo son certos.



MEMORIA DO PROXECTO DE INVESTIGACIÓN

DOCUMENTO Nº 1-B

32618861P

cod_web: 388

IN825B

AIB 1

TÍTULO
MODELOS DE PREDICION CON DATOS FUNCIONAIS. APLICACIÓNS MEDIOAMBIENTAIS E FINANCEIRAS.
INVESTIGADOR/A PRINCIPAL
WENCESLAO GONZALEZ MANTEIGA
ORGANISMO
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA
ENDEREZO
RUA LOPE GOMEZ DE MARZOA S/N CAMPUS SUR
PALABRAS CLAVE
SERIES DE TEMPO; DATOS FUNCIONAIS; POLUCION; ACTIVOS FINANCEIROS
CONTIDO DO PROXECTO: RESUMO
<p>Un dos campos de maior interese da Estatística é a análise de series de tempo. Durante os últimos anos desenvolvéronse numerosas metodoloxías enfocadas a modelar e predicir o comportamento destas series, tanto parametricamente coma non parametricamente. Xurdiron así diversos modelos adicados á estimación da tendencia (desde os clásicos ARMA, ARIMA, ata redes neuronais ou modelos aditivos) e da volatilidade (comezando polos ARCH, GARCH ata os modelos non paramétricos máis recentes). Ademais, deseñáronse diversas técnicas de refinamento co obxectivo de mellorar os resultados dados por ditos modelos, como por exemplo o Boosting e o Bagging.</p> <p>No presente proxecto preténdese trasladar as ferramentas estatísticas anteriormente expostas a unha das áreas máis novidosas dos últimos tempos: os datos funcionais. Esta extensión presenta un grande atractivo tanto desde o punto de vista académico, xa que a investigación no eido funcional está nos seus inicios, coma desde o punto de vista empírico, posto que na actualidade en diferentes campos das ciencias aplicadas (enxeñería medioambiental, química, biométrica, medicina, econometría...) os datos recolectados son curvas. Os métodos obtidos ao longo deste proxecto poderán ser aplicados en contextos como o Medio Ambiente (predición de niveis de diversos contaminantes atmosféricos) e a Economía (predición de prezos de activos financeiros). Estas aplicacións nacen da 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. e Caixa Galicia.</p>

CONTIDO DO PROXECTO: ESTADO DA ARTE DA INVESTIGACIÓN PROPOSTA

A análise de series de tempo, tanto dende o punto de vista teórico como empírico, constitúe unha parte fundamental dos estudos estatísticos da maioría das ciencias aplicadas. Dado o carácter aleatorio de numerosas variables de interese desenvóléronse diversas técnicas estatísticas, 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 aos puramente autorregresivos (AR) e aos modelos de medias móbiles (MA) (Wei, 1990; Brockwell e Davis, 1987). Estes modelos foron xeneralizados ao caso multidimensional dando lugar aos modelos VARMA (Lütkepohl, 2005). Á vista das limitacións destes modelos á hora de modelizar a volatilidade, introducíronse nos últimos anos modelos de series de tempo non lineais (Fan e Yao, 2003), como por exemplo os modelos de tipo ARCH e GARCH e as súas correspondentes extensións ao caso multidimensional (Bauwens, et al., 2003). Por exemplo, os modelos VEC, BEKK, DVEC, F-GARCH ou CCC son algúns dos modelos GARCH multivariantes.

Por outra banda, tamén se desenvolveron modelos non paramétricos para a estimación da tendencia e da volatilidade como os modelos aditivos (Hastie e Tibshirani, 1990), modelos baseados en redes neuronais (Tsay, 2002) ou a estimación tipo núcleo.

Recentemente xurdiron técnicas destinadas a mellorar as predicións proporcionadas polos modelos citados anteriormente como o Boosting e o Bagging (Hastie, Tibshirani e Friedman, 2001). O Boosting é un potente método nacido no eido das técnicas de aprendizaxe. A motivación desta técnica era obter un clasificador como combinación de outros máis débiles. A pesar de ser orixinalmente deseñada para problemas de clasificación, foi estendida á regresión e algoritmos de aprendizaxe en xeral, xa que reduce o erro de predición. Pola súa parte o Bagging está baseado na metodoloxía Bootstrap e busca diminuír a varianza deste erro mediante a remostraxe das predicións.

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. A metodoloxía funcional aínda está en desenvolvemento pero xa se teñen presentado avances en numerosos eidos como no caso dos modelos de regresión con datos funcionais con procedementos próximos ao suavizado spline (Ramsay e Silverman, 2005) ou ben mediante a extensión da estimación tipo núcleo (Ferraty e Vieu, 2006).

A preocupación pola degradación do medio natural e a actual lexislación sobre niveis de polución, fan imprescindible a utilización de metodoloxía estatística para a predición de concentracións de distintas sustancias no medio. En particular, para a predición de niveis de SO₂ no entorno da Central Térmica de As Pontes utilizáronse modelos semiparamétricos (García Jurado, et al., 1995), modelos parcialmente lineais (Prada Sánchez, et al., 2000), modelos baseados en redes neuronais (Fernández de Castro, et al., 2003), modelos con resposta binaria (Roca Pardiñas, et al., 2004), modelos aditivos (Cotos Yáñez, et al., 2004), modelos funcionais (Fernández de Castro, et al., 2005).

Un dos campos de aplicación das máis novidosas técnicas estatísticas está a ser o das finanzas (o que incluso xerou unha nova disciplina, a econometría financeira, que ten coma obxectivo primordial a análise e contraste de datos e modelos financeiros). Neste ámbito a estatística pode proporcionar ferramentas de decisión en áreas como a predición de prezos, selección de carteira ou estimación do Values at Risk. Unha das cuestións centrais no estudo dos mercados financeiros é a preditibilidade do prezo dos activos ben traballando con datos de baixa frecuencia (por exemplo, de tipo mensual) ben con datos de alta frecuencia (por exemplo de tipo horario). Proba do interese que suscitou a econometría financeira é a recente aparición de libros especializados na aplicación de técnicas estatísticas ás series financeiras (Chan, 2002; Dacorogna, Gençay, Müller e Pictet, 2001; Gouriéroux, 1997; Gouriéroux e Jasiak, 2001; Tsay, 2002).

Por razóns de espazo omítese as referencias dos libros e algúns dos artigos que se atopan referenciados na Experiencia do Equipo.

Fernández de Castro, B., et al. (2003), Prediction of SO₂ level using neural networks. *Journal of the air and waste management association* 53, 532-538.

García Jurado, I., et al. (1995), Predicting using Box-Jenkins, nonparametric and bootstrap techniques. *Technometrics* 37 303-310.

Prada Sánchez, J.M., et al. (2000), Prediction of SO₂ pollution incidents near a power station using partially linear models and a historical matrix of predictor-response vectors. *Environmetrics* 11, 209-225.

32618861P

cod_web: 388

AIB 1 IN825B

CONTIDO DO PROXECTO: OBJECTIVOS DO PROXECTO

Neste proxecto existen dúas vías de investigación principais. A continuación detállanse os obxectivos concretos a acadar en cada unha de elas:

1. Elaboración de metodoloxía dentro do campo dos datos funcionais.

- 1.1. Estudo do modelo de regresión lineal con covariables funcionais e resposta real.
- 1.2. Estudo do modelo de regresión lineal con covariables funcionais e resposta funcional.
- 1.3. Deseño dunha nova clase de estimadores a partir de estimadores de mínima distancia.
- 1.4. Elaboración de tests para a selección de modelos.
- 1.5. Extensión a datos dependentes.
- 1.6. Estudos de simulación dos modelos anteriores.
- 1.7. Aplicación dos modelos estudados en 1.1. e 1.2. á predición de series medioambientais e financeiras.

2. Elaboración de metodoloxía dentro do campo dos modelos de series temporais e das técnicas de aprendizaxe e refinamento.

- 2.1. Revisión de modelos paramétricos e nonparamétricos en series de tempo.
- 2.2. Revisión de modelos aditivos.
- 2.3. Revisión de técnicas de aprendizaxe e refinamento.
- 2.4. Extensión destas ideas ao contexto de datos funcionais.
- 2.5. Aplicacións dos modelos estudados nos apartados anteriores á predición de series medioambientais e financeiras.

3. Elaboración de software que permita a transferencia tecnolóxica da nova metodoloxía á Central Térmica de As Pontes.

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

32618861P

cod_web: 388

AIB 1 IN825B

DOCUMENTO ENGADIDO A SOLICITUDE CO SEU RESUMO DIXITAL SHA1:

78FAC90BB73CDF5ECCBC60072797CF199B4EB3F1 - Diagrama.pdf

CONTIDO DO PROXECTO: INTERESE PARA O AVANCE DO COÑECEMENTO E DA SOCIEDADE

A temática dos datos funcionais é de recente aparición polo que hai un amplo abanico de liñas de investigación aínda sen explorar. Por iso este proxecto, que pretende 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. 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.

En particular poden ser empregados para tratar series financeiras e poder predicir, a partir da curva de prezos de activos financeiros dun día, os prezos futuros (predición puntual), superacións de certos valores fixos (predición binaria) ou incluso, obter curvas completas de ditos prezos (predición funcional). As mesmas técnicas permitirán facer estudos similares con datos de tipo medioambiental como os niveis de distintos contaminantes atmosféricos (por exemplo, dióxido de xofre, óxidos de nitróxeno, ...)

32618861P

cod_web: 388

AIB 1 IN825B

CONTIDO DO PROXECTO: 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 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) e, aproveitando que o departamento pertence ao programa CONSOLIDER, incluírase 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 activos financeiros, deseños de carteiras,... o que favorecerá a transferencia tecnolóxica de todos estes avances teóricos ao mundo empresarial (Posibles beneficiarios: ENDESA e Caixa Galicia).

32618861P

cod_web: 388

AIB 1 IN825B

VIABILIDADE DO PROXECTO: METODOLOXÍA

No que ao desenvolvemento matemático se refire, a metodoloxía a seguir será a habitual nestes casos empregando tanto ferramentas estatísticas como probabilísticas: 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). 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 cinco membros do grupo reuniranse semanalmente. Por outra banda, as viaxes que unha das bolseiras realiza semanalmente 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 e Caixa Galicia favorecerá dita transferencia ao eido financeiro.

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 currículos 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.

32618861P

cod_web: 388

AIB 1 IN825B

DOCUMENTO ENGADIDO A SOLICITUDE CO SEU RESUMO DIXITAL SHA1:

98018878940E26EAE4DB7FA31922A15DC519F67F - CartaEndesa_Diag2.pdf

VIABILIDADE DO PROXECTO: PLAN DE TRABAJO (ETAPAS, RECURSOS ASIGNADOS, CUMPRIMENTO OBTIVOS, CRONOGRAMA,...)

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 aos datos funcionais. 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 e 2.1 a 2.3.

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 técnicas de aprendizaxe e refinamento ao campo funcional. Isto corresponde ao cumprimento dos obxectivos 1.3 a 1.6 e 2.4.

Terceiro ano.

Aplicaranse os resultados obtidos durante os dous anos anteriores a datos medioambientais e financeiros.

Ademais o software construído favorecerá a transferencia tecnolóxica a empresas como ENDESA e Caixa Galicia. Isto corresponde ao cumprimento dos obxectivos 1.7 e 2.5.

Todos os membros do equipo participarán activamente nas tarefas anteriormente mencionadas. Ademais previsiblemente incorporaranse ao grupo, en canto obteña o DEA, Abelardo Monsalves, actualmente estudante do primeiro ano do doutoramento de Estatística e Investigación Operativa.

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

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

32618861P

cod_web: 388

AIB 1 IN825B

DOCUMENTO ENGADIDO A SOLICITUDE CO SEU RESUMO DIXITAL SHA1:

858006D3743671566D0A6476255C4A28C273EEA4 - CRONOGRAMA.pdf

PRESUPUESTO: EXPLICACIÓN E XUSTIFICACIÓN DE CADA UN DOS CONCEPTOS DE GASTO QUE CONFORMAN O CUSTO TOTAL DO PROXECTO

A continuación xustificanse os gastos que aparecen na Memoria Económica.

MATERIAL INVENTARIABLE OU BIBLIOGRÁFICO.

1. Actualización de 5 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 aboan 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.

32618861P

cod_web: 388

AIB 1 IN825B

CAPACIDADE E EXPERIENCIA DO EQUIPO: EXPERIENCIA DO EQUIPO NA DIRECCIÓN DOUTROS PROXECTOS DE I+D+i. LOGROS OBTIDOS NOS ÚLTIMOS CINCO ANOS

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 omítese a participación en proxectos con empresas, capítulos de libros ou proceedings, e numerosas aportacións en distintos congresos de rango nacional e internacional. 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 NO_x levels: location and scale estimation and outlier detection". *Computational Statistics* Aceptado para publicación.
- 3)Cadarsó Suárez, Carmen María e González-Manteiga, W. (2007 (in press)). "Statistics in Biomedical Research". *ARBOR*.
- 4)Roca Pardiñas, Javier; Cadarsó 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)Cadarsó 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; Cadarsó 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)Matías 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.; Martínez Miranda, María 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 Cadarsó Suárez, C. M. (2004). "Predicting binary time series of SO₂ using generalized additive models with unknown link function". *Environmetrics* 15, p. 729-742.

32618861P

cod_web: 388

AIB 1 IN825B

CAPACIDADE E EXPERIENCIA DO EQUIPO: INSTALACIÓNS, INSTRUMENTOS E TÉCNICAS DISPONIBLES PARA A REALIZACIÓN DO PROXECTO.

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).

32618861P

cod_web: 388

AIB 1 IN825B

CAPACIDADE E EXPERIENCIA DO EQUIPO: OUTROS MEDIOS NECESARIOS NON DISPONIBLES.

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.

32618861P

cod_web: 388

AIB 1 IN825B

XUSTIFICACIÓN DA SUBVENCIÓN SOLICITADA NO DOCUMENTO 1-A (en euros)

GASTOS DE PERSOAL CONTRATADO		
SUBTOTAL		0
MATERIAL INVENTARIABLE OU BIBLIOGRÁFICO (XUSTIFICACIÓN DETALLADA)		
ACTUALIZACION DE 5 ORDENADORES PERSOAIS E DO EQUIPO INFORMATICO DA CENTRAL TERMICA DE AS PONTES		9.500
MATRIZ SAS SUN STORAGE TEK 2530 876GB		9.000
BIBLIOGRAFIA		3.000
TRADUCION E EDICION DE ARTICOS		3.000
SUBTOTAL		24.500

(en euros)

Documento 1B-17

[illegible]

32618861P

cod_web: 388

AIB 1 IN825B

(en euros)

Documento 1B-18

OUTROS GASTOS (XUSTIFICACIÓN DETALLADA).

SUBTOTAL

0

(Euros)

IMPORTE DO PROXECTO (I)

59.200

CUSTOS INDIRECTOS (15% de (I)) (II)

8.880

SUBVENCIÓN SOLICITADA (I) + (II)

68.080

O/A investigador/a principal, que asina máis abaixo, declara coñecer as normas específicas da presente convocatoria

SANTIAGO DE COMPOSTELA , 19 de xuño de 2007

Asdo.: WENCESLAO GONZALEZ MANTEIGA

DECLARACIÓN

D./D.^a WENCESLAO GONZALEZ MANTEIGA, co DNI n.º 32618861P,
investigador principal do equipo solicitante dunha axuda para o proxecto
MODELOS DE PREDICION CON DATOS FUNCIONAIS. APLICACIÓNS MEDIOAMBIENTAIS E FINANCEIRAS.

DECLARA:

☒ Que non ten solicitado nin, en consecuencia, concedida ningunha axuda para o mesmo fin ao que se refire a presente solicitude, por ningunha das Administracións Públicas competentes ou polos seus organismos ou sociedades.

☐ Que ten solicitado e, no seu caso, concedidas as seguintes axudas para a mesma finalidade:

Denominación do organismo, sociedade ou entidade pública á que lle foi solicitada a axuda	Datos de (1)		Importe	
	Solicitude	Concesión(2)	Solicitado	Concedido(3)

(1) Especificar o D.O.G., B.O.E., etc. no que foi convocada e/ou resolta a axuda solicitada.

(2) No caso de estar pendente a resolución dalgunha solicitude, indicárase "PENDENTE" nesta columna.

(3) Nas solicitudes denegadas farase constar "0 euros" nesta columna.

☒ Que se compromete a utilizar a lingua galega en todas as relacións que, como consecuencia deste procedemento, manteña coa administración autonómica, de acordo co establecido no artigo 6.3 da Lei 3/1983, do 15 de xuño, de normalización lingüística.

De conformidade cos apartados 3 e 4 do artigo 13 da Lei 4/2006, do 30 de xuño, de transparencia e de boas prácticas na Administración pública galega, e co previsto no Decreto 132/2006, do 27 de xullo, polo que se regulan os rexistros públicos creados nos artigos 44 e 45 da Lei 7/2005, do 29 de decembro, de orzamentos xerais da Comunidade Autónoma de Galicia para o ano 2006, AUTORIZO EXPRESAMENTE á Consellería á publicación no Diario Oficial de Galicia, na súa páxina web oficial e nos rexistros públicos correspondentes, dos datos de carácter persoal que esixen as referidas normas, referidos tanto a información sobre as axudas concedidas como as sancións, que como consecuencia delas puidesen impoñerse.

SANTIAGO DE COMPOSTELA , 19 de xuño de 2007

(sinatura)