

Group Description	
Title of Research Group:	(RG-MATH-LVT-Lisboa-6-951) Statistical Modelling in Environmental and Life Sciences
Principal Investigator:	Maria Antonia Conceição Abrantes Amaral Turkman
Main Scientific Domain:	Matemática
Group Host Institution:	Faculdade de Ciências - Universidade de Lisboa

Funding, source, dates
Funding, source, dates
<p>ALL THE REFERRED AMOUNTS INCLUDE OVERHEADS TO PAY FFCUL</p> <p>1. FCT Base for Research Group Title: (RG-MATH-LVT-Lisboa-6-951) - Statistical Modelling in Environmental and Life Sciences</p> <p>Funding:</p> <p>corresponding to 2008:</p> <p>date: 29 Abril 2008 / 31021.28 Eur</p> <p>date: 26 Novembro 2008 / 33347.87 Eur</p> <p>Corresponding to 2006:</p> <p>date: 29 Dezembro 2008 / 27888.25 Eur</p> <p>2. Projects: The following Project has Principal Investigator from the Group:</p> <p>(RG-MATH-LVT-Lisboa-6-951) - Statistical Modelling in Environmental and Life Sciences</p> <p>FCT/ PTDC/MAT/64353/2006</p> <p>MEGA – STATISTICAL METHODS IN GENETIC AND ENVIRONMENT</p> <p>Funding Entity: PTDC - FCT</p> <p>Total Award Ammount: 60 000 Eur</p> <p>Prime Contractor: Universidade de Lisboa - Fundação da F.C.U.L.</p> <p>Period Covered: from September of 2007 to September of 2010</p> <p>Coordinator: Prof. K.F.Turkman</p>

Objectives & Achievements
Objectives
<p>The main objective of this research group is to make contributions on specific probability and statistical issues in modelling environmental and biological data, as well as, to create a group with expertise in analyzing data coming from these fields. To achieve this goal we seek collaboration with researchers from the environmental and biological sciences, not only with the objective to gather data to apply the methodology developed, but as well to help them in solving their own problems. Another objective is to continue the Ph.D. programme on the main research themes, namely Hierarchical Bayesian Modelling, Spatial-temporal Modelling, Statistics in Genetics. Together with applied research we have as an objective to continue developing fundamental research in the areas of interest, namely on</p> <ul style="list-style-type: none"> • Extremes of Random Fields; • Some theoretical and practical aspects of continuous -and discrete- time extremes • State-space models. • Marked point processes • Hierarchical Bayesian spatiotemporal analysis • Analyses of categorical data with missingness in responses and factors

Objectives & Achievements

- Latent class models
- Model selection and model adequacy from a Bayesian perspective
- Elicitation of prior information in presence of nuisance parameters
- Generalized additive models with flexible link in survival analysis
- Survival analysis and longitudinal models
- Screening methods in Supervised classification
- Disease mapping
- Ecological studies
- Modelling infectious diseases
- Stochastic modelling of dynamic systems
- Syndromic surveillance
- Bayesian networks
- Hidden Markov models
- Statistical problems in Cellular Biology and Genetics
- Microarray data analysis
- Bayesian analysis of allelic penetrance models for complex binary traits.

Main Achievements

- Study of the relationship between max-stable models for spatial extreme data and Bayesian hierarchical models (BHM).
- Quantification of the regional risk of large wildfire sizes, by fitting a Generalized Pareto distribution to excesses over a suitably chosen high threshold. Spatio-temporal variations are introduced into the model through model parameters with suitably chosen link functions. The inference on these models are carried using Bayesian Hierarchical Models and Markov chain Monte Carlo methods.
- Bayesian methods to model jointly the probability of ignition and fire sizes in Australia. The models and conclusions bring improvements on the results reported by Russell-Smith et al. (2007) based on a similar data set.
- Continuation of the study of the evolution of the human resources of the Faculty of Sciences of Lisbon and projections for 2020, using Bayesian hierarchical methods.
- Work focused on some theoretical issues that can occur in the applications in the genome analysis, namely: (a) The problem of multiple hypothesis testing, when Fisher's exact test or Chi-Square test are used to compare frequencies of rare mutations in two independent groups at several positions in the genome of hepatitis C virus (HCV). (b) The power of some normality tests under a mixture of Gaussian distributions.
- Development of statistical models to genetics and epidemiological data, namely (a) Development of statistical prognostic models in survival context, (b) Longitudinal Data Analysis of cardiovascular biomarkers and cohort data, (c) Bayesian hierarchical model applied to mapping morbidity data, (d) Generalized Linear Models in exploratory models investigating associated factors to health outcomes
- Implementation of a background correction method for RIP-chip (RNA immunoprecipitation on chip) inspired in a non-specific hybridization method used for pre-processing data from ChIP-Chip technology. A new non-parametric statistical method for RIP-Chip arrays is also being developed.
- Some normalization and gene selection methods were used on data obtained by Gama-Carvalho et al. (2006), with the goal of identifying mRNA molecules associated with the splicing factors PTB and U2AF65, in HeLa cells.
- Application of some Bayesian methods (BAM, B-statistics and BGX) for selecting the enriched RNA in the immuno precipitated samples and comparison with some classical methods (RankProd, Fold Change, t-statistic and SAM).
- Some insights on the assessment of the predictive value of diagnostic tests with application to some respiratory tests in Amyotrophic Lateral Sclerosis.
- A comparative study of Bayesian and Non-Bayesian classification methods to identify differential expressed genes.
- Some new advances in supervised classification methods, namely the use of optimal screening methods in gene profile classification. In this work, it was proposed a new Bayesian classification methodology ("screening" classifier) based entirely on marker gene pairs expression monitoring by DNA microarrays. The classifier is more general than several others described in literature and results were very satisfactory. A script in R was written to help the implementation of the procedure
- Continuation of the collaborative work on the application of advanced statistical methods to (i) genetic susceptibility to Rheumatoid Arthritis, to (ii) the study of the predictive value of some respiratory tests in Amyotrophic Lateral Sclerosis and (iii) the analysis of the longitudinal variations of CD40L at the onset of acute myocardial infarction and along the recovery period.
- The study of the prevalence of tuberculosis in New York City using counting processes with two change-points
- Bayesian analysis of allelic penetrance models for complex binary traits
- Deprivation analysis based on Bayesian latent class models
- Hierarchical Bayesian spatiotemporal analysis of revascularization odds using smoothing splines

Objectives & Achievements

- Aneuploidy and high S-phase as biomarkers of poor clinical outcome in poorly differentiated and anaplastic thyroid carcinoma Most of this work was presented as invited or contributed papers in International Conferences, published as extended abstract; or submitted for publication.

Group Productivity

Publications in peer review Journals

1. Achcar, J. A., Martinez, E. Z., Ruffino-Netto, A., Paulino, C. D. and Soares, P. (2008). A statistical model investigating the prevalence of tuberculosis in New York City using counting processes with two change-points. *Epidemiology and Infection*, 136, 1599-1605.
2. Silva, G.L., Dean, C.B, Niyonsenga, T. and Vanasse, A. (2008). Hierarchical Bayesian spatiotemporal analysis of revascularization odds using smoothing splines. *Statistics in Medicine*, 27, 2381-2401. Link: <http://www3.interscience.wiley.com/journal/116331945/abstract?CRETRY=1&SRETRY=0>
3. Pinto, A.E., Silva, G.L., Banito, A., Leite, V. and Soares, J. (2008). Aneuploidy and high S-phase as biomarkers of poor clinical outcome in poorly differentiated and anaplastic thyroid carcinoma. *Oncology Reports*, 20, 913-919.
a. Link: <http://www.spandidos-publications.com/or/20/4/913>
4. Gil, S., Sepúlveda, N., Albina, E., Leitão, A., Martins, C. (2008). The low virulent African swine fever virus (ASFV/NH/P68) induces enhanced expression and production of relevant regulatory cytokines (IFN α , TNF α and IL12p40) on porcine macrophages in comparison to the highly virulent ASFV/L60. *Arch. Virol.* 153, 1845-1854.
5. Souto-Carneiro, M. M., Fritsch, R., Sepúlveda, N., Lagareiro, M. J., Morgado, N., Longo, N. S. and Lipsky, P. E. (2008). The NF-kappaB canonical pathway is involved in the control of the exonucleolytic processing of coding ends during V(D)J recombination. *The Journal of Immunology* 180, 1040-1049.
6. Pereira, S.A., Caixas, U., Branco, T., Germano, I., Lamprea, F., Papoila, A.L. & Monteiro, E.C. (2008). Efavirenz concentrations in HIV-infected patients with and without viral hepatitis. *British Journal of Clinical Pharmacology*. 66:4, 551-555.
7. Martins, S.O.M., Serra e Silva, P. Papoila, A.L., Caramona, M., Foppe van Mil, J.W. & Cabrita, J. (2008). Assessment of global cardiovascular risk and risk factors in Portugal according to the SCORE $\text{\textcircled{R}}$ model. *J Public Health*. 16:361-367.
8. Acioli-Santos, B., Sebastiana, M., Pessoa, F., Sousa, L., Figueiredo, A., Fortes, A.M., Baldé, A., Maia, L.C. and Pais, M.S. (2008). Fungal transcript pattern during the preinfection stage (12h) of ectomycorrhiza formed between *Pisolithus tinctorius* and *Castanea sativa* roots, identified using cDNA microarrays. *Current Microbiology*. (DOI:10.1007/s00284-008-9253-2)
9. Figueiredo, A., Fortes, A.M., Ferreira, S., Sebastiana, M., Choi, Y.H., Sousa, L., Acioli-Santos, B., Pessoa, F., Verpoorte, R. and Pais, M.S. (2008). Transcriptional and metabolic profiling of grape (*Vitis vinifera* L.) leaves unravel possible innate resistance against pathogenic fungi. *Journal of Experimental Botany* 59(12): 3371-3381. (DOI:10.1093/jxb/ern187)
10. Fortes, A.M., Santos, F., Choi, H.Y., Silva, M.S., Figueiredo, A., Sousa, L., Pessoa, F., Santos, B.A., Sebastiana, M., Palme, K., Malho, R., Verpoorte, R. and Pais, M.S. (2008). Organogenic nodule development in hop (*Humulus lupulus* L.): transcript and metabolic responses. *BMC Genomics* 9:445. (DOI:10.1186/1471-2164-9-445)
11. Gonçalves, L. and Amaral Turkman, MA (2008) Triangular and Trapezoidal Distributions: Applications in the Genome Analysis. *Journal of Statistical Theory and Practice*, vol2, p. 45-54
12. Antunes, M. and Sousa, L. (2008). Bayesian classification and non-Bayesian label estimation via EM algorithm to identify differentially expressed genes: a comparative study. *Biometrical Journal* 50(5): 824-836. (DOI: 10.1002/bimj.200710468)
13. Lopes C, Andreozzi VL, Ramos E, Sá Carvalho M. (2008) Modelling over week patterns of alcohol consumption Alcohol and alcoholism (Oxford, Oxfordshire) 43(2):215-22

Other publications International

1. Ramos, S., Amaral Turkman, M.A. Antunes, M. (2008). Optimal Screening Methods in Gene Expression Profiles Classification. In *Proceedings in Computational Statistics 2008*. Physica – Verlag Heidelberg, vol II, part III, 134-144
2. Fonseca Mde J, Andreozzi VL, Faerstein E, Chor D, Carvalho MS. (2008) Alternatives in modeling of body mass index as a continuous response variable and relevance of residual analysis. *Reports in Public Health*. 2008 Feb;24(2):473-8.
3. Geórgia Chalfun, Rosane Reis de Mello, Maria Virginia Peixoto Dutra, Valeska Andreozzi, Kátia Silveira da Silva. (2008) Risk Factor for Respiratory Morbidity at 12 to 36 months in Very Low Birth Weight Premature Infants Previously submitted to a Public Neonatal Intensive Care Unit. *Reports in Public Health*. Accepted for publication in December 2008
4. Teles, J.; Gonçalves, L. (2008) Power of normality tests under a mixture of Gaussian Distributions: a simulation Study. In *Proceedings of 23rd International Workshop on Statistical Modelling* (P. Eilers, Editor), pp.400-413
5. Gonçalves Pereira, M; Saboga Nunes, L; Papoila, A; Maia, T.; Loureiro, I; Xavier, M. (2008). Sense of Coherence in Caregivers of the severely mentally ill – Pathways for Mental Health Promotion. In *Proceedings of the 16th European Conference on Public Health (EUPHA)*, Lisbon, 2008 (European Journal of Public Health, vol 18 Suppl 1, p 209).

Other publications National

Group Productivity

Books:

Carvalho, M. L. e Natário, I. (2008). *Análise de Dados Espaciais*. Sociedade Portuguesa de Estatística. (ISBN: 978-972-8890-18-6)

Papers in Portuguese Periodicals (with peer review)

Mário Gentil Belard Silvano Barata e Ana Luísa Papoila (2008). Consumo de drogas na marinha portuguesa: estudo de prevalência. *Toxicodependências*, vol. 14, 3, 59-65.

Patrícia Napoleão, Mafalda Selas, Alexandra Toste, Antónia Turkman, Valeska Andreozzi, Ana Maria Viegas-Crespo, Teresa Pinheiro, Rui Cruz Ferreira. (2008) Serial Changes of Oxidized Low-density Lipoprotein Associated with Culprit Vessel in ST-Elevation Myocardial Infarction – A Promising Marker? *Revista Portuguesa de Cardiologia*. Artigo accepted November 2008

Chapters of books (with peer review)

Poleto, F., Singer, J. e Paulino, C.D. (2008). Análise log-linear em dados categorizados estratificados com respostas omissas. Em *Estatística da Teoria à Prática* M M Hill, et al (eds), 437-450, SPE.

Sepúlveda, N., Paulino, C.D. e Carneiro, J. (2008). Diversidade de linfócitos T reguladores e efetores. Em *Estatística da Teoria à Prática* M M Hill, et al (eds), 513-523, SPE.

Machado, C., Paulino, C.D. e Nunes, F. (2008). Análise multidimensional da pobreza em Portugal com base em modelos bayesianos de classes latentes. Em *Estatística da Teoria à Prática* M M Hill, et al (eds), 277-287, SPE

Carina Silva-Fortes, Emiliano Barreto-Hernandez, Lisete Sousa, Maria Antónia Amaral Turkman, Margarida Gama-Carvalho (2008) Análise por microarrays de complexos ribonucleoproteicos: normalização e selecção de genes com associação diferencial. In *Estatística da Teoria à Prática*, M M Hill, et al (eds), 537-549, SPE

Papers in proceedings

Goes, M. & Natário, I. & Oliveira, M. M. & Bonito, J (2008). "Educação Factores Risco Institucionalização Idoso: um Estudo no Distrito Beja." *Actas do II Congresso Nacional de Educação para a Saúde*. Centro de Investigação em Educação e Psicologia da Universidade de Évora

Master and Ph.D. thesis completed

Ph.D. thesis:

1. Elisete Maria Rodrigues Correia, February 2008 'O Papel de Distribuições Condicionamente Especificadas em Estatística Bayesiana': Supervised by: Maria Antónia Amaral Turkman

Master thesis :

1. Ana Luísa de Sousa Lourenço Quitério, 2008, "Modelos de regressão dinâmica na revisão das Séries do Inquérito ao Emprego", supervised by Lucília Carrvalho.

2. Sónia Maria Sousa Freitas, 2008 "Associação entre os Polimorfismos Genéticos e a Hipertensão Arterial na População da Região Autónoma da Madeira", supervised by Fernanda Oliveira

3. Rute Gomes Velosa Vieira, 2008 "Predição de Variantes Genéticas do VIH-1" supervised by Lisete Sousa

4. Encarnação, F.; O Modelo de classes Latentes com Restrições Aplicado à Análise do Desempenho de Testes de Diagnóstico; Dissertação de Mestrado, Instituto Superior Técnico; Lisboa; 2008; 133 p. (concluída; Orientador: M. Rosário Oliveira, co-orientador Luzia Gonçalves).

5. Ferreira, C (2008) Influência da Poluição do Ar na Ocorrência de algumas Doenças Respiratórias: Relações Temporais e Espaciais. Relatório de estágio, Universidade de Évora. (concluído; Orientador: Dulce Gomes, co-orientador Luzia Gonçalves).

Organization of conferences

1- Análise de dados espaciais – Short course for the XVI Congresso da Sociedade Portuguesa de Estatística. Lucília Carvalho and Isabel Natário

2- Advanced course: Hierarchical Modelling and Analysis for Space-Time Data, by Professor Alan Gelfand, from Duke University, USA. March 2008, CEAUL, Faculdade de Ciências da Universidade de Lisboa, Portugal. This 2 days course was organized by KF Turkman was attended by more than 50 participants.

3- Curso Prático e Avançado em Análise de Sobrevivência: Teoria e Aplicações em Saúde. Lecturers: Valeska Andreozzi and Marília Sá Carvalho (Fiocruz, Brasil). July 2008 CEAUL, Faculdade de Ciências da Universidade de Lisboa, Portugal. This 3 days course was organized by Valeska Andreozzi.

4- EQS 2008, Estatística e Qualidade na Saúde. 20-21 November 2008, Escola Superior de Tecnologias da Saúde. Carina Silva Fortes was the coordinator of the organizing committee and Antónia Amaral Turkman, a member of the Scientific Committee.

5- COMPSTAT 2008, Porto, Portugal, on August 24th-29th 2008, Antónia Amaral Turkman was a member of the scientific committee

Internationalization

Having regard to the Decision No 234/2008 of the European Parliament and of the Council of 11 March 2008, Maria Lucília Salema e Carvalho was appointed by the Council of the European Union of 4 November 2008, as the member representing the Council in the European Statistical Advisory Committee (ESAC).

"responsible for the transition of the Labour Force Survey from mode CAPI to CATI".

1 Invited talks in International Conferences

K.F. Turkman was an invited speaker in the following conferences:

Group Productivity

International Workshop on Applied Probability, Université de Technologie de Compiègne, France, July 7-10 2008, with the following talk: Turkman, KF, Amaral Turkman, MA, Mendes, J (2008) Models for extremes of spatio-temporal data, Proceedings of the International Workshop on Applied Probability. edited by Julien Chiquet, Joseph Glaz, Nikolaos Limnios, and Pascal Moyal (abstract, available on CD, is in C3 b).

PCI 2008 Problems of Cybernetics and Informatics, 10-12 September 2008, Baku, Azerbaijan, with the following talk: Anderson, CW and Turkman, KF, (2008) Extremes of Continuous Processes and Their Discrete Versions. PCI 2008 Problems of Cybernetics and Informatics, 10-12 September 2008, Baku, Azerbaijan. (in C4 b, the summary published in book of abstracts)

M.A. Amaral Turkman was an invited speaker in the following conferences:

Third Workshop on Statistics, Mathematics, and Computation First Portuguese-Polish Workshop on Biometry, 21st-22nd July, 2008, Communications Museum Lisbon, with the following talk: Amaral Turkman, MA, Faria S. and Carvalho, M.(2008) Predictive tools in the assessment of diagnostic tests.

Plenary session, II Iberian Mathematical Meeting October 3-5, 2008, Badajoz, Spain with the talk: Amaral Turkman, MA, Faria S. and Carvalho, M. Predictive tools in the assessment of diagnostic tests.

PCI 2008 Problems of Cybernetics and Informatics, 10-12 September 2008, Baku, Azerbaijan with the talk: Amaral Turkman, M:A: and Correia, E. (2008) Bayes Inference on the ratio between two proportions.

Daniel Paulino was an invited speaker at the II Iberian Mathematical Meeting October 3-5, 2008, Badajoz, Spain, with the talk: Bayesian analysis of allelic penetrance models for complex binary traits.

Government/Organization contract research

Valeska Andreozzi:

Period : Since June 2007 Function : Biostatistician Institution : Portuguese Ministry of Health Unit : Office of High Commissioner for Health Department : Dept. of Statistical Analysis

Job summary : Monitoring and evaluation of the National Health Plan 2004-2010

Job description: To analyse a set of health indicators in primary intervention areas for which the National Health Plan has defined targets for 2010, two works are worth of mentioned. The first one compared the impact of different statistical models for long term health indicators prediction using short time series and taking into account the epidemiological interpretation of the models assumptions. The second work applied frailty survival models to determine the underlying risk for 30-day in-hospital mortality among patients admitted with acute myocardial infarction in Portugal using information provided by the hospitals diagnoses related group database (DRG)

Future Research

Objectives

1. Objectives (3000 ca.)

1 To prepare the 2nd edition of the textbooks

Bayesian Statistics (D. Paulino and MA Amaral Turkman);

Análise de Sobrevida: Teoria e Aplicações em Saúde. (V. Andreozzi)

2. Finish the book on "Survival Analysis" which is a topic for the short course for SPE conference (C. Rocha and AL Papoila)

3. Work on the books:

Non-linear time series (Feridun Turkman);

Modelagem Estatística: Aplicações à Saúde" (V. Andreozzi, M. Antunes and collaborators from FIOCRUZ)

4 To get several submitted papers published on the following themes :

Missing Categorical Data;

Supervised classification;

Spatio-temporal extremes

Identification of plant genes

Pre-processing Optimization of RNA Immunoprecipitation Microarray Data

5 To finish the final version of the papers "

A new parametric Bayesian approach for cytogenetic dosimetry";

Patterns of HIV Portuguese interments: a hierarchical survival model

A Bayesian Approach to Gene Expression Profile Classification

Identification of plant genes involved on the initial contact between ectomycorrhizal symbionts.

Future Research

Sample size for estimating a binomial proportion: comparison of different methods.

Spatial and temporal extremes of wildfire sizes in Portugal (1984-2004)

Hierarchical space-time models for high dimensioned fire data.

Pre-processing Optimization of RNA Immunoprecipitation Microarray Data.

Recurrent wheezing after Respiratory Syncytial Virus bronchiolitis: a longitudinal study.

Bayesian Approach to Gene Expression Profile Classification

Mercury, cadmium and lead bioaccumulation in black scabbard fish, *Aphanopus carbo* Lowe, 1839: a case study from mainland Portugal, Azores and Madeira Archipelagos

Obtaining a Mortality Baseline free of influenza epidemics effects using models with no covariates

6 To prepare versions for submission of papers on "

Modelling and mapping of disease mortality and incidence rates by age-groups over public health regions in Canada"

Joint analysis of counts and severity with zero inflated longitudinal data

Comparing diagnostic tests with missing data

A product-multinomial framework for categorical data analysis with missing responses

Missing data mechanisms and their implications on the analysis of categorical data

The use of Bayesian Hierarchical Models for supervised classification - an application to Rheumatoid Arthritis

Modelling Strategies for Longitudinal Studies

Predictive Tools in the Assessment of Diagnostic Tests

Assessing the impact of ecological bias in ecological studies

Integration of syndromic surveillance data for outbreak detection of infectious diseases

Relations between mercury concentrations in different tissues of black scabbard fish

Approaches for predicting mortality in intensive care units

7 To finish the work concerning Sepúlveda's PhD. thesis "How is the T-cell repertoire shaped?" and submit a paper on Inferences based on models for T lymphocyte repertoire diversity in Immunology.

To finish the work concerning Leonel Vicente's PhD. Thesis on "Human Resources of the FCUL: a Bayesian hierarchical model".

Funding, source, dates

ALL THE REFERRED AMOUNTS INCLUDE OVERHEADS TO PAY FFCUL

1. FCT Base for Research Group (RG-MATH-LVT-Lisboa-6-951) Statistical Modelling in Environmental and Life Sciences :

A. Funding left to 2009 from 2008:

Due to the delay and late transferring dates from FCT (at the end of the year), the Group was not able to schedule a convenient execution of a percentage from the budget of previous year, namely: around 77975.36 Eur.

B. Funding from 2009:

Already transferred the first semester amount from FCT on the 3rd April 2009: 38169.25 Eur

Expected Second semester amount from FCT : similar to first semester

2. Projects: Partially source funding is still available during 2009 and 2010:

FCT/ PTDC/MAT/64353/2006

MEGA – STATISTICAL METHODS IN GENETIC AND ENVIRONMENT

Funding Entity: PTDC - FCT

Total Award Amount: 60 000 Eur

Prime Contractor: Universidade de Lisboa - Fundação da F.C.U.L.

Period Covered: from September of 2007 to September of 2010

Coordinator: Prof. K.F.Turkman