

The Fourth International Workshop on Environmental Applications of Machine Learning EAML 2004

Proceedings

Bled, Slovenia
September 27 – October 1, 2004

Edited by

Sašo Džeroski
Bernard Ženko
and
Marko Debeljak

Jožef Stefan Institute
Ljubljana, Slovenia

CIP-Kataložni zapis o publikaciji
Narodna in univerzitetna knjižnica, Ljubljana

004.85:574(063)(082)
004.85:504(063)(082)

INTERNATIONAL Workshop on Environmental Applications of Machine Learning
(4 ; 2004 ; Bled)

Proceedings / The Fourth International Workshop on Environmental Applications of Machine Learning – EAML 2004, Bled, Slovenia, September 27–October 1, 2004 ; edited by Sašo Džeroski, Bernard Ženko and Marko Debeljak. – Ljubljana : Jožef Stefan Institute, 2004

ISBN 961-6303-59-7
1. Džeroski, Sašo

215423744

Editors:
Sašo Džeroski
Bernard Ženko
Marko Debeljak

Production, design, and type provided by the conference organization

Cover image, photograph of Bled, Slovenia, courtesy of Marjan Smerke,
Jožef Stefan Institute, Ljubljana, Slovenia

Publisher:
Jožef Stefan Institute,
Jamova 39,
SI-1000 Ljubljana, Slovenia
<http://www.ijs.si>

Copyright © 2004 by Jožef Stefan Institute
All rights reserved

Printed in Slovenia by GRAFIKO, d.o.o., Grosuplje
Ljubljana, September 2004

Foreword

The Fourth International Workshop on Environmental Applications of Machine Learning (EAML) and the Fourth European Conference on Ecological Modelling (ECEM) were organized jointly in Bled, Slovenia, during the week of September 27 - October 1, 2004. The aim of these events was to bring together researchers from the areas of ecology, ecological modelling and environmental sciences, on one hand, and the areas of data analysis, data mining and machine learning, on the other. In many respects, they were treated as a single event, namely ECEM/EAML 2004.

Ecological modelling is concerned with the development and use of mathematical, computer and simulation models of ecosystems. It is a relatively young scientific discipline, which is rapidly gaining importance, especially because of the use and usefulness of ecological models for the management of natural resources. The European Conference on Ecological Modelling is the premiere European scientific event in the area of ecological modelling and regularly attracts an international audience.

As environmental concerns grow and information technology develops, more and more data on the different aspects (physical, chemical, biological, ecological) of the environment are gathered. There is an increasing need to analyse the collected environmental data for different purposes, which include the support for environmental management decisions. The International Workshop on Environmental Applications of Machine Learning provided a forum for presenting recent advances in applying machine learning and data mining techniques for the analysis of environmental data.

EAML 2004 covered all topics related to the application of data mining and machine learning methods to environmental data. An indicative, but non-exhaustive list of topics is given below:

- Analysis of environmental data with:
 - computational scientific discovery;
 - decision and regression trees;
 - evolutionary computing, e.g., genetic algorithms and programming;
 - statistical learning, e.g., kernel methods and SVMs;
 - neural networks, e.g., MLPs and SOMs;
 - probabilistic methods, e.g., Bayesian networks;
 - relational learning methods;
 - rule induction methods.

- Data mining and machine learning for:
 - modelling of different types of ecosystems, e.g., agricultural ecosystems, aquatic ecosystems, grassland ecosystems, forest ecosystems;
 - modelling different aspects of ecosystems/ecological processes, e.g., biodiversity changes, habitat suitability, population dynamics;
 - analysis of different types of environmental data, e.g., monitoring data (air/soil/water samples; chemical, physical, biological), remote sensing data (e.g., on atmosphere, geology, vegetation), spatial data (e.g., GIS data on land cover), temporal data (e.g., time series data on pollution levels);
 - various environmental applications of machine learning, e.g., for decision support in environmental management, for earthquake prediction, for environmental risk assessment, in environmental epidemiology, in meteorological/atmospheric sciences, in predictive ecotoxicology.

A total of 110 abstracts were submitted and 90 were accepted after a review process. Each submitted abstract was sent to three members of the Program Committee for review. The Program Committee members themselves did the majority of reviews, assisted by a few additional reviewers appointed by the Committee members. Authors of accepted abstracts (for oral or poster presentation) were invited to submit full versions of their papers. These were reviewed separately by the members of the Program Committee and other reviewers for inclusion in a special issue of the journal *Ecological Modelling* which was to be published after the conference.

The program of EAML 2004 consisted of three invited talks by Joseph C. Coughlan (NASA Ames, USA), Cesare Furlanello (ITC-irst, Trento, Italy), and Jacqueline McGlade (European Environment Agency), as well as oral and poster presentations of accepted contributions.

Many people contributed in various ways to the EAML 2004 workshop. We would especially like to thank:

- The authors of submitted contributions who made the workshop possible by presenting their work;
- The three invited speakers: Joseph C. Coughlan, Cesare Furlanello, and Jacqueline McGlade;
- The Advisory Committee for their suggestions regarding the Program Committee, invited speakers and encouraging remarks;
- The members of the Program Committee for their efforts in evaluating the submitted abstracts and papers;
- The members of the Organizing Committee;
- The sponsors for their generous support;

- The management and staff of the Albatros Congress Turist Agency, Bled, and the Center for Knowledge Transfer in Information Technologies, Jožef Stefan Institute, Ljubljana for their support;
- The Jožef Stefan Institute for providing the organizational infrastructure.

Sašo Džeroski, Bernard Ženko and Marko Debeljak
EAML 2004 Program Co-chairs

Contents

Foreword	iii
Acknowledgements	xi
Workshop Chairs	xi
Organizing Committee	xi
International Advisory Committee	xi
Program Committee	xii
Organizational Support	xii
Invited Talks	1
An overview of ecological modelling and machine learning research within the U.S. National Aeronautics and Space Administration <i>Joseph C. Coughlan</i>	3
GIS-based predictive models for ecology <i>Cesare Furlanello (joint work with S. Merler, S. Menegon, M. Neteler, S. Fontanari, R. Blažek, A. Rizzoli, and C. Chemini)</i>	5
Spatial assessments of Europe’s environment <i>Jacqueline McGlade</i>	7
Contributed Abstracts	9
Accounting for different sources of temporal and spatial autocorrelation in non- linear habitat preference models <i>Aarts, G., Matthiopoulos, J., McConnell, B., MacKenzie, M., and Fedak, M.</i>	11
Computational revision of ecological process models <i>Asgharbeygi, N., Langley, P., Bay, S., and Arrigo, K.</i>	13
Modelling spatial distribution of Croatian marine habitats <i>Bakran-Petricioli, T., AntoniĆ, O., Bukovec, D., Petricioli, D., Janeković, I., Križan, J., Kušan, V., and Dujmović, S.</i>	15
Model selection in demographic time series using VC-bounds <i>Corani, G., and Gatto, M.</i>	17

Characterizing vertical forest stands structure using data mining methods <i>Debeljak, M., and Babič, J.</i>	19
Using multiobjective classification to model communities of soil microarthropods <i>Demšar, D., Džeroski, S., Krogh, P. H., and Larsen, T.</i>	21
The use of data mining for the monitoring and control of anaerobic waste water treatment plants <i>Gallop, J. R., Dixon, M., Lambert, S. C., Lardon, L., and Steyer, J.-P.</i>	23
Grid-based data analysis of air pollution data <i>Ghanem, M., Guo, Y., Hassard, J., Osmond, M., and Richards, M.</i>	25
A new multiobjective strategy to support model selection for environmental modelling <i>Giustolisi, O., Savić, D. A., and Doglioni, A.</i>	27
Using wavelets for the classification of hyperspectral images <i>Jasso, H. and Shin, P.</i>	29
Application of machine learning methods to palaeoecological data <i>Jeraj, M., Džeroski, S., Todorovski, L., and Debeljak, M.</i>	31
Analysis and spatial modelling of red deer's (<i>cervus elaphus</i> l.) winter and annual habitat in the dinaric forest of south-western Slovenia with decision trees in a raster gis environment <i>Jerina, K., and Adamič, M.</i>	33
Using neural networks and GIS to predict the spatial occurrence of freshwater fish and decapods <i>Joy, M. K., and Death, R. G.</i>	35
Habitat mapping using machine learning-extended kernel-based reclassification of an Ikonos satellite image <i>Kobler, A., Džeroski, S., and Keramitsoglou, I.</i>	37
A new synergetic paradigm including machine learning and deterministic components for environmental numerical modeling <i>Krasnopolksy, V. M., and Fox-Rabinovitz, M. S.</i>	39
CARROTAGE: A software for mining land-use data <i>Le Ber, F., Mari, J.-F., Benoit, M., Mignolet, C., and Schott, C.</i>	41
Learning to predict channel stability using biogeomorphic features <i>Moret, S. L., Langford, W. T., and Margineantu, D. D.</i>	43
Artificial neural networks and time series modelling of a forested watershed <i>Nour, M. H., Smith, D. W., El-Din, M. G., and Prepas, E. E.</i>	45
Using regression trees to estimate surface water runoff and soil erosion for range-lands <i>Pachepsky, Y., Pierson, F., and Weltz, M.</i>	47
Neural network modelling for the analysis of forcings/temperatures relationships at different scales in the climate system <i>Pasini, A.</i>	49
Radon forecasting in the low atmosphere by neural network modelling and estimation of the stable layer depth <i>Pasini, A., Ameli, F., and Lore, M.</i>	51

Analysis of radon concentrations in Slovenian thermal waters for earthquake prediction <i>Popit, A., Todorovski, L., Zmazek, B., Vaupotič, J., Džeroski, S., and Kobal, I.</i>	53
Improvement of emergency vaccination strategies against rabies in red fox (<i>Vulpes vulpes</i>) populations using a combination of Cellular Automata and Evolutionary Algorithms <i>Selhorst, T.</i>	55
Sparse regression for analyzing the development of foliar nutrient concentrations in coniferous trees <i>Sulkava, M., and Tikka, J.</i>	57
Using the expert model PERPEST to translate measured and predicted pesticide exposure data into ecological risks <i>Van den Brink, P., Brown, C. D., and Dubus, I. G.</i>	59
Segmentation of paleoecological spatio-temporal count data <i>Vasko, K., Toivonen, H., and Korhola, A.</i>	61
Modelling lake Glumsoe with Q ² learning <i>Vladušič, D., Kompare, B., and Bratko, I.</i>	63
Q ² Prediction of ozone concentrations <i>Žabkar, J., Žabkar, R., Vladušič, D., Čemas, D., Šuc, D., and Bratko, I.</i>	65
Automatic construction of concept hierarchies: The case of foliage-dwelling spiders <i>Žnidaršič, M., Jakulin, A., and Džeroski, S.</i>	67
Author Index	69

Acknowledgements

We acknowledge the support of the following institutions:

Jožef Stefan Institute, Ljubljana, Slovenia

Nova Gorica Polytechnic, Nova Gorica, Slovenia

Slovenian Ministry for Education, Science and Sport, Ljubljana, Slovenia

ISEM, The International Society for Ecological Modelling

KD-net, The Knowledge Discovery Network

PASCAL – Pattern Analysis, Statistical Modelling and Computational Learning,
The network of excellence

Workshop Chairs

Sašo Džeroski

Bernard Ženko

Marko Debeljak

Organizing Committee

Tina Anžič

Jana Babič

Mili Bauer

Damjan Demšar

Peter Ljubič

International Advisory Committee

Ivan Bratko

Thomas G. Dietterich

Cesare Furlanello

Pat Langley

Sovan Lek

Donato Malerba

Friedrich A. Recknagel

Marko Slokar

William J. Walley

Program Committee

Nataša Atanasova	Li Bai-Lian
Vladan Babović	Horst Malchow
Marko Bohanec	Donato Malerba
Ivan Bratko	Joao Carlos Marques
Broder Breckling	Robert I. McKay
Bert Bredeweg	Felix Mueller
Ulises Cortes	Søren N. Nielsen
Marko Debeljak	Mark O'Connor
Thomas G. Dietterich	Luca Palmeri
Yannis Dimopoulos	Bernard C. Patten
Sašo Džeroski	Friedrich A. Recknagel
Nicola Fohrer	Miquel Sanchez-Marre
Tony Fountain	Dragan Savić
Cesare Furlanello	Mark Schwabacher
Jasna Grbović	Masahiko Sekine
Francois Guerrin	Cosimo Solidoro
Sven E. Jørgensen	Yuri M. Svirezhev
Michio J. Kishi	Ljupčo Todorovski
Andrej Kobler	Sergio Ulgiati
Alexander S. Komarov	Tanja Urbančič
Boris Kompare	Alexey Voinov
Branko Kontić	Hans Voss
Vipin Kumar	Monica Wachowicz
Tetsuya Kusuda	William J. Walley
Bill Langford	Peter A. Whigham
Pat Langley	Franz Wotawa
Tarzan Legović	Bernard Ženko
Sovan Lek	

Organizational Support

The Albatros Congress Tourist Agency, Bled, Slovenia
Center for Knowledge Transfer in Information Technologies,
Jožef Stefan Institute, Ljubljana, Slovenia