MANUELA PAVAN

Department of Environmental Sciences
UNIVERSITY OF MILANO-BICOCCA

Correspondence	**Home Address**
Department of Environmental Sciences University of Milano-Bicocca P.za della Scienza, 1 20126 Milano (Italy) Phone: +39 02 64482818 Fax. +39 02 64482890 E-mail: manuela.pavan@unimib.it	Strada al lago, 18 S.Felice Segrate 20090, Milano, Italy Phone: +39 02 7530495

CURRICULUM VITAE

ACADEMIC POSITION

- Present: **University assistant researcher**
  University of Milano-Bicocca, Department of Environmental Sciences.

EDUCATION

- 2000 - 2003: Department of Environmental Sciences, University of Milano-Bicocca, Italy
  Ph.D. in **Chemical Sciences**. Thesis: "Total and Partial Ranking Methods in Chemical Sciences".
- February - June 2002: **Stage** at the School of Mathematical and Physical Sciences, James Cook University of North Queensland, Townsville, Australia.
- 1999 - 2000: University of Insubria, Varese, Italy
  QSAR fellowship on "Studies of biodegradability of organic compounds by quantitative structure-activity relationships".
- 1994 - 1999: University of Milano-Bicocca, Milano, Itlay
  Degree in Environmental Sciences (*Summa cum laude*). Thesis: "Studi sulla predizione della mutagenicità di ammine aromatiche mediante metodi chemiometrici".

FIELDS OF INTEREST

- *Primary*: Chemometrics, QSAR, Molecular Descriptors, Chemometric Software Development, Priority setting, Total and Partial Ranking methods, Multicriteria Decision Making Strategies.
- *Secondary:* Experimental Design, Sensory analysis.

SOFTWARE DEVELOPED

- Co-author of **DRAGON** software. DRAGON is a software package for the calculation of molecular descriptors developed by Milano Chemometrics and QSAR Research Group.
- Co-author of **MOBYDIGS** software. MOBYDIGS is a software package developed by Milano Chemometrics and QSAR Research Group for the calculation of regression models by using genetic algorithms for variable selection to obtain an optimal subset of predictive models.

- Co-author of **RANA** software: RANA (Ranking ANalysis Alghoritms) is a software package for total and partial ranking procedures.

TEACHING

- March 2004: Teaching-Assistant in the Second Italian School on Molecular Descriptors, Talete srl, University of Milano-Bicocca, Milano, Italy.
- January 2004: Teaching-Assistant for practical lessons of the course "Chemiometria" in the Department of Environmental Sciences, University of Milano-Bicocca, Milano, Italy.
- October 2003: Teaching-Assistant for practical lessons of the course "I fondamenti della pianificazione sperimentale" at the Lever Fabergé, Casalpusterlengo, Italy.
- November 2002: Teaching-Assistant for practical lessons of the course "Laboratorio di Integrazione III" in the Department of Environmental Sciences, University of Milano-Bicocca, Milano, Italy.
- June 2001: Teaching-Assistant for the graduate course of "Modelística superior aplicada a problemas tecnológicos y de gestíon ambiental" in the Ecuadorian Master "Programa de Postgrado de Maestría en Gestíon Tecnologíca", Universidad Del Azuay (UDA), Cuenca (Ecuador).
- June 2000: Teaching-Assistant for the graduate course of "Diseno Experimental y Optimizacion de Procesos" in the Ecuadorian Master "Programa de Postgrado de Maestría en Gestíon Tecnologíca", Universidad Del Azuay (UDA), Cuenca (Ecuador).

ACADEMIC ACTIVITY

- **Poster** on "Structural – Toxicity mode of action similarity analysis by Kohonen artificial neural networks (K-ANN)" P.Gramatica, V.Consonni, M.Pavan, *14th Annual Meeting of SETAC Europe*, Prague (Czech Republic), 18 - 22 April 2004.
- **Poster** on "Comparison of several different fitness functions in regression models" A.Mauri, D.Ballabio, V.Consonni, M.Pavan, and R.Todeschini. *Colloquium Chemometricum Mediterraneum*, Ustica, Italy, 25 - 27 June 2003.
- **Poster** on "New fitness functions to avoid bad regression models in variable subset selection by using Genetic Algorithms" R.Todeschini, V.Consonni and M.Pavan. *Colloquium Chemometricum Mediterraneum*, Ustica, Italy, 25 - 27 June 2003
- **Poster** on "Data mining by a partial ranking strategy (Hasse Diagram Technique, HDT)" M.Pavan, A.Mauri, V.Consonni, D.Ballabio and R.Todeschini. *Colloquium Chemometricum Mediterraneum*, Ustica, Italy, 25 - 27 June 2003.
- **Poster** on "Development of order ranking models by Genetic Algorithm Variable Subset Selection (GA-VSS). M.Pavan, V.Consonni and R.Todeschini. *Colloquium Chemometricum Mediterraneum*, Ustica, Italy, 25 - 27 June 2003.
- **Poster** on "The development of a ranking system based on welfare parameters for commercial dairy farms" M.V.Tosi, S.Mattiello, M.Pavan, E.Canali, C.Carenzi. *37th International Congress of the International Society for Applied Ethology*, Abano Terme, Italy, 24 - 28 June 2003.
- **Poster** on "Chemometric methodologies for the modelling of the heterogeneous chemical toxicity: dataset representativity as the absolute essential" P.Gramatica, V.Consonni, M.Pavan, P.Pilutti, E.Papa *13th Annual Meeting of SETAC Europe,* Hamburg (Germany), 27 April - 1 May 2003.
- **Poster** on "Utilizo di tecniche chemiometriche nel confronto tra i profili sensoriali di prodotti cosmetici" M.Ferrero, M.Pavan and R.Todeschini. *XXIX Congresso Nazionale della Società Italiana di Chimica e* Scienze *Cosmetologiche*, Milan, Italy, 22 November 2002.

- **Poster** on "Distance Measure between Models: a Basic Tool for Model Similarity/Diversity Analysis." R.Todeschini, V.Consonni, M.Pavan. *Kemometria, Hungarian Chemometric Workshop*, Tata, Ungaria, 29 September-1 October 2002.

- **Poster** on "New fitness functions to avoid bad regression models in variable subset selection by Genetic Algorithms." R.Todeschini, V.Consonni, A.Mauri, M.Pavan. *14th European Symposium on quantitative structure-activity relationships*, Bournemouth, UK, 8-13 September 2002.

- **Poster** on "Distance Measure between Models: a Basic Tool for Model Similarity/Diversity Analysis." R.Todeschini, V.Consonni, M.Pavan. *14th European Symposium on quantitative structure-activity relationships*, Bournemouth, UK, 8-13 September 2002.

- **Poster** on "Heavy metal and lignin structure variation in some conifer in Valdaosta (Alps in north-west Italy) from 1940." M.Orlandi, M.Pelfini, M.Pavan, M.P.Colombini. *X hungarian-Italian Symposium on Spectrochemistry: Trace substance in the biosphere Eger*, 1-5 October 2001.

- **Poster** on "Metodi chemiometrici per l'identificazione del legante utilizzato per la preparazione dei supporti lignei nella pittura italiana del XII – XVI secolo." M. P.Colombini, R.Todeschini, V.Consonni, M.Pavan, R. Lleti, M. C.Ortiz. *XVI National Congress of Analitical Chemistry*, Portonovo, Ancona, Italy 24-28 September 2001.

- **Poster** on "Modelling and prediction of pollutant global persistence by theoretical molecular descriptors" P.Gramatica, E.Papa and M.Pavan. *11th Annual Meeting of SETAC Europe*, Madrid (Spain), 6-10 May 2001.

- **Poster** on "Bridging the Gap between Science and Regulation of Chemical Mixtures: the "BEAM" EU Research Project". T.Backaus, R.Altenburger, A.Arrhenius, H.Blanck, F.Consolaro, M.Faust, A.Finizio, P.Gramatica, M.Grothe, M.Junghans, W.Meyer, M.Pavan, M.Scholze, R.Todeschini, M.Vighi, S.Villa, H.Walter and L.H.Grimme. *11th Annual Meeting of SETAC Europe*, Madrid (Spain), 6-10 May 2001.

- **Poster** on "Physico-chemical properties modelling for environmental pollutants." F.Consolaro, P. Gramatica and M.Pavan. *9th International Workshop on Quantitative Structure Activity Relationships in Environmental Sciences – QSAR 2000*, Bourgas (Bulgaria), 16-20 September 2000.

- **Poster** on "QSAR modelling of the biodegradability by holistic molecular descriptors." P.Gramatica, M.Pavan, F.Consolaro, V.Consonni and R.Todeschini. *9th International Workshop on Quantitative Structure Activity Relationships in Environmental Sciences – QSAR 2000*, Bourgas (Bulgaria), 16-20 September 2000.

- **Poster** on "Organic pollutants environmental fate: modelling and prediction of global persistence by molecular descriptors." P.Gramatica, F.Consolaro and M.Pavan. *9th International Workshop on Quantitative Structure Activity Relationships in Environmental Sciences – QSAR 2000*, Bourgas (Bulgaria), 16-20 September 2000.

- **Poster** on "Risk Assessment of Chemical Mixtures in the AquaticEnvironment: an outline of the BEAM Project" L.H.Grimme, R.Altenburger, A.Arrhenius, T.Backaus, H.Blanck, F.Consolaro, M.Faust, A.Finizio, P.Gramatica, M.Grothe, M.Junghans, M.Pavan, M.Scholze, R.Todeschini, M.Vighi, S.Villa and H. Walter. *EurOCEAN 2000 The European Conference on Marine Science and Ocean Technology*, Hamburg (Germany), 29 August-2 September 2000.

- **Poster** on "QSAR modelling of the aromatic amines mutagenicity by genetic algorithm-variable subset selection." M.Pavan, P.Gramatica, F.Consolaro, V.Consonni and R.Todeschini. *9th International Workshop on Quantitative Structure Activity Relationships in Environmental Sciences – QSAR 2000*, Bourgas (Bulgaria), 16-20 September 2000.

- **Poster** on "Mutagenicity of aromatic amines: modelling, prediction and classification by new molecular descriptors." M.Pavan and P.Gramatica. *3rd SETAC World Congress and 10th Annual Meeting of SETAC-Europe*, Brighton (UK), 21-25 May 2000.

- **Poster** on "POP environmental fate: modelling and prediction of half-life ranges and octanol/air partition coefficient (Koa) by molecular descriptors." P.Gramatica, F.Consolaro, S.Pozzi and M.Pavan. *3rd SETAC World Congress and 10th Annual Meeting of SETAC-Europe,* Brighton (UK), 21-25 May 2000.

SCHOOLS

- Course of Statistical Methods, at the School of Mathematical and Physical Sciences, James Cook University of North Queensland, Townsville, Australia February-June 2002.

- English for Academic Purposes Course at the James Cook University English Language Centre, Townsville, Australia February-May 2002.

- VIII[a] Italian School of chemometrics, Italian Society of chemometrics, Genova, 25-29 September 2000.

- First Italian School on Molecular Descriptors, Italian Society of chemometrics, Gargnano del Garda, September28-30, 1998.

PAPERS

- **Detecting 'bad' regression models: multicriteria fitness functions in regression analysis**. (2004).

  R.Todeschini, V.Consonni, A.Mauri and M.Pavan. *Analityca Chimica Acta*, **515**, 199-208.

  Abstract: In variable selection it is often not just one model but a population of models that is provided. All these models are optimal according to the specific fitness function used by the variable selection algorithm. However, if validation is carried out, some of them later turn out to lack predictive ability. Models with good fitting but no predictive ability are sometimes chance correlations and often show some pathological features such as multicollinearity, overfitting, and inclusion of random variables. This problem is well known and of the utmost importance.
  The present paper proposes new constraints for model fitness functions, the aim being to recognize linear regression models with pathology. These constraints have been thought of in order to face the following problems: model instability due to outliers and influential objects; predictor multicollinearity; redundancy in explanatory variables; overfitting due to chance factors.
  In model searching, our proposal is to mazimize the Q2 statistics with the new constraints so as to obtain a final optimal population of models. Computations on the Selwood data set are reported to illustrate the use of these constraints in model searching.

- **New indices for analysing partial ranking diagrams**. (2004).

  M.Pavan and R.Todeschini. *Analytica Chimica Acta,* **515**, 167-181.

  *Abstract***:** Interest is growing in decision making strategies and several techniques are now available. The assessment of priorities is a typical premise before final decisions are taken. Total and Partial order ranking (POR) strategies, which from a mathematical point of view are based on elementary methods of Discrete Mathematics, appear as an attractive and simple tool to asses priorities. Despite to the well-known total ranking strategies, which are scalar methods combining the different criteria values into a global index which always ranks elements in an ordered sequence, the partial order ranking is a vectorial approach which recognises that not all the elements can be directly compared with all the others. In fact when many criteria are considered, contradictions in the ranking are bound to exist and the higher the number of criteria, the higher the probability that contradictions in the ranking occur. The Hasse diagram technique (HDT) is a very useful tool to perform partial order ranking. The results of the partial order ranking are visualised in a diagram, called Hasse diagram. Incomparable elements are located at the same geometrical height and as high as possible in the diagram, thus incomparable elements are arranged in levels. The quality of a ranking procedure has to be evaluated by a deep analysis and by several indices, i.e. scalar functions that describe features of an ordered set and allow comparison among different rankings. For this purpose, new indices for ranking analysis are here proposed, compared with the ones found in literature and tested on theoretical examples and on real data.

- **MobyDigs: software for regression and classification models by genetic algorithms**. (2004)

  R.Todeschini, V.Consonni, A.Mauri and M.Pavan in "Nature-inspired methods in chemometrics: genetic algorithms and artificial neural networks" (R.Leardi Ed.), Chapter 5, p.141-167 - Elsevier, 2004.

- **New fitness functions to avoid bad regression models in variable subset selection by Genetic Algorithms.** (2004).

R.Todeschini, V.Consonni, A.Mauri and M.Pavan in "Designing Drigs and Crop Protectants: processes, problems and solutions." (M. Ford, D. Livingstone, J. Deardean, H. van de Waterbeemd Ed.), Chapter 5, p.323-325 – Blakwell, 2004.

- **Data mining by total ranking methods: a case study on optimisation of the "pulp and bleaching" process in the paper industry** (2004)

  M.Pavan, R.Todeschini and M.Orlandi. *Chemometrics and Intelligent Laboratory Systems,* submitted.

  Abstract: Total order ranking methods are multicriteria decision making techniques used for the ranking of various alternatives on the basis of more than one criterion. The criteria, which are the standards by which the elements of the system are judged are not always in agreement, they can be conflicting, motivating the need to find an overall optimum that can deviate from the optima of one or more of the single criteria. Total order ranking methods are based on an aggregation of the criteria in a scalar function, i.e. an order or ranking index, which allow to sort elements according to its numerical value. Several evaluation methods which define a ranking parameter generating a total order ranking have been proposed in the literature. Four total order ranking methods are here described: Desirability functions, Utility functions, Dominance functions and Absolute Reference method and the proposed Relative Reference method. These methods have been compared to each other by applying them to a decision making problem in the paper industry. Various bleaching processes have been analysed and compared on the basis of multiple criteria, the aim being to find out best bleaching process among the ones proposed in the last years as alternative to chlorine bleaching process which is of high environmental impact due to the potential for chlorinated dioxin production.

- **Total ranking models by Genetic Algorithms Variable Subset Selection (GA-VSS) approach for environmental priority settings** (2004)

  M.Pavan, A.Mauri and R.Todeschini. *Analytical and Bioanalytical Chemistry*, in press.

  Abstract: Total order ranking (TOR) strategies, which from a mathematical point of view are based on elementary methods of Discrete Mathematics, appear as an attractive and simple tool to perform data analysis. Moreover order ranking strategies seem to be a very useful tool not only to perform data exploration but also to develop order ranking models, being a possible alternative to conventional QSAR methods. In fact, when data material is characterised by uncertainties, order methods can be used as alternative to statistical methods such as multi-linear regression (MLR), since they do not require specific functional relationship between the independent variables and the dependent variables (responses). A ranking models is a relationship between a set of dependent attributes, experimentally investigated, and a set of independent attributes, i.e. model attributes, which are calculated attributes. As in regression and classification models the variable selection model is one of the main step to find predictive models. In the present work, the Genetic Algorithm (GA-VSS) approach is proposed as the variable selection method to search for the best ranking models within a wide set of variables. The models based on the selected subsets of variables are compared with the experimental ranking and evaluated by the Spearman's rank index. A case study application is presented on a total order ranking model developed for polychlorinated biphenyl compounds, which have been analysed according to some of their physico-chemical properties which play an important role on their environmental impact.

- **Partial Ranking Models by Genetic Algorithm Variable Subset Selection (GA-VSS) approach for environmental priority settings** (2004)

  M.Pavan, V.Consonni and R.Todeschini. *MATCH Communications in Mathematical and in Computer Chemistry,* in press.

  Abstract: Partial order ranking (POR) strategies, which from a mathematical point of view are based on elementary methods of Discrete Mathematics, appear as an attractive and simple tool to perform data analysis. Moreover order ranking strategies seem to be a very useful tool not only to perform data exploration but also to develop order ranking models, being a possible alternative to conventional QSAR methods. In fact, when data material is characterised by uncertainties, order methods can be used as alternative to statistical methods such as multi-linear regression (MLR), since they do not require specific functional relationship between the independent variables and the dependent variables (responses).
  A ranking models is a relationship between a set of dependent attributes, experimentally investigated, and a set of independent attributes, i.e. model attributes, which are calculated attributes. As in regression and classification models the variable selection model is one of the main step to find predictive models. In the present work, the Genetic Algorithm (GA-VSS) approach is proposed as the variable selection method to search for the best ranking models within a wide set of variables. The models based on the selected subsets of variables are compared with the

experimental ranking and evaluated by a set of similarity indices. A case study application is presented on a partial order ranking model developed for 23 chemicals selected as active ingredients used in agricultural practice and analysed according to their toxicity on *Scenedesum vacuolatus*.

- **Study of anaerobic and aerobic degradation of different non ionic surfactants**. (2003).

  V. Mezzanotte, F. Castiglioni, R. Todeschini, M. Pavan. *Bioresource Technology*, **87**, 87-91.

  Abstract: Six alcohol ethoxylates (C5E2, C6E4, C7E4, C8E2, C8E4, C10E4) and two fatty acid esthers were tested at lab-scale for degradation in anaerobic and aerobic conditions and OUR. Anaerobic removal of C5E2, C6E4 and C7E4 improved with the increasing of the number of ethoxy groups and with the decreasing of the length of the alkyl chain. Their aerobic removal was also remarkable but lower than anaerobic one. C8E2, C8E4 and C10E4 were adsorbed on sludge but not degraded in anaerobic conditions while they were more efficiently removed in aerobiosis. The fatty acid esthers were removed to an intermediate level between the two alcohol ethoxylates groups in both anaerobiosis and aerobiosis. The measured OURs confirmed the different behaviour of the three groups of compounds.

- **Prediction of Aromatic Amines Mutagenicity by Theoretical Molecular Descriptors**. (2003).

  P.Gramatica, V.Consonni and M.Pavan, *SAR and QSAR in Environmental Research*, **14**, 237-250.

  Abstract: In the present research the mutagenicity data (Ames tests TA98 and TA100) for various aromatic and heteroaromatic amines, a data set extensively studied by other QSAR-authors, have been modelled by a wide set of theoretical molecular descriptors using linear multivariate regression (MLR) and Genetic Algorithm – Variable Subset Selection (GA-VSS). The models have been calculated on a subset of compounds selected by a D-optimal Experimental Design. Moreover, they have been validated by both internal and external validation procedures showing satisfactory predictive performance. The models proposed here can be useful in predicting data and setting a testing priority for those compounds for which experimental data are not available or are not yet synthesised.

- **The BEAM-project: prediction and assessment of mixture toxicities in the aquatic environment.** (2003)

  T.Backhaus, R.Altenburger, Å.Arrhenius, H.Blanck, M.Faust, A.Finizio, P.Gramatica, M.Grothe, M.Junghans, W.Meyer, M.Pavan, T.Porspring, M.Scholze, R.Todeschini, M.Vighi, H.Walter and L.H.Grimme, *Continental Shelf Research,* **23** (17-19), 1757-1769.

  Abstract Freshwater and marine ecosystems are exposed to various multi-component mixtures of pollutants. Nevertheless, most ecotoxicological research and chemicals regulation focus on hazard and exposure assessment of individual substances only, the problem of chemical mixtures in the environment is ignored to a large extent. In contrast, the assessment of combination effects has a long tradition in pharmacology, where mixtures of chemicals are specifically designed to develop new products, e.g. human and veterinary drugs or agricultural and non-agricultural pesticides. In this area, two concepts are frequently used and are thought to describe fundamental relationships between single substance and mixture effects: Independent Action (Response Addition) and Concentration Addition. The question, to what extent these concepts may also be applied in an ecotoxicological and regulatory context may be considered a research topic of major importance, as the concepts would allow to make use of already existing single substance toxicity data for the predictive assessment of mixture toxicities. Two critical knowledge gaps are identified: (a) There is a lack of environmental realism, as a huge part of our current knowledge about the applicability of the concepts is restricted to artificial situations with respect to mixture composition or biological effect assessment. (b) The knowledge on what exactly is needed for using the concepts as tools for the predictive mixture toxicity assessment is insufficient. Both gaps seriously hamper the necessary, scientifically sound consideration of mixture toxicities in a regulatory context. In this paper, the two concepts will be briefly introduced, the necessity of considering the toxicities of chemical mixtures in the environment will be demonstrated and the applicability of Independent Action and Concentration Addition as tools for the prediction and assessment of mixture toxicities will be discussed. An overview of the specific aims and approaches of the BEAM project to fill in the identified knowledge gaps is given and first results are outlined.

- **Structure/Response Correlations and Similarity/Diversity Analysis by GETAWAY Descriptors. Part 1. Theory of the Novel 3D Molecular Descriptors**. (2002)

  V.Consonni, R.Todeschini and M.Pavan. *J.Chem.Inf.Comput.Sci*., **42**, 3, 682-692.

Abstract: Novel molecular descriptors based on a leverage matrix similar to that defined in statistics and usually used for regression diagnostics are presented. This leverage matrix, called *Molecular Influence Matrix* (MIM), is here proposed as a new molecular representation easily calculated from the spatial coordinates of the molecule atoms in a chosen conformation.
The proposed molecular descriptors are called *GETAWAY* (GEometry, Topology and Atom-Weights AssemblY) as they try to match 3D-molecular geometry provided by the molecular influence matrix and atom relatedness by molecular topology, with chemical information by using different atomic weightings (atomic mass, polarizability, van der Waals volume, and electronegativity, together with unit weights).
A first set of molecular descriptors, called *H-GETAWAY*, is derived by using only the information provided by the molecular influence matrix, while a second set, called *R-GETAWAY*, combines this information with geometric interatomic distances in the molecule.
The prediction ability in structure-property correlations of the new descriptors was tested by analyzing regressions of these descriptors for selected properties of octanes.

- **Structure/Response correlations and Similarity/Diversity Analysis by GETAWAY Descriptors. Part 2. Applications of the Novel 3D Molecular Descriptors to some QSAR/QSPR Studies**. (2002).

  R.Todeschini, V.Consonni, M.Pavan and P.Gramatica. *J.Chem.Inf.Comput.Sci*., **42**, 3, 693-705.

  Abstract: In a previous paper the theory of the new molecular descriptors called *GETAWAY* (GEometry, Topology and Atom-Weights AssemblY) was explained. These descriptors have been proposed with the aim of matching 3D-molecular geometry, atom relatedness and chemical information.
  In this paper prediction ability in structure-property correlations of GETAWAY descriptors has been tested extensively by analyzing the regressions of these descriptors for selected properties of some reference compound classes. Moreover, the general performance of the new descriptors in QSAR/QSPR has been evaluated with respect to other well-known sets of molecular descriptors.

- **Heavy metals variations in some conifers in Valdaosta (western italian Alps) from the 1930 to 2000**. (2002).

  M.Orlandi, M.Pelfini, M.Pavan, M.Santilli, M.Colombini. *Microchemical Journal*. **73**, Issues 1-2, 237-244.

  Abstract: The high mountain environment is very sensitive to the climatic and ecological variations, that are registered in several natural archives as glaciers or plants. Trees in particular modify their growth, habitus, ring features and spatial distribution in relation with climate and environmental changes. Heavy metals variation in tree rings of *Larix decidua* have been determined to asses whether this arboreal species can be used as bio-geochemical tracers of heavy metal pollution of the alpine environment.

- **A Distance Measure between Models: a Tool for Similarity/Diversity Analysis of Model Populations.** (2002).

  R.Todeschini, V.Consonni, and M.Pavan. *Chemometrics and Intelligent Laboratory System*, Elsevier Science, **70**, 55-61.

  Abstract: In many research fields there is, nowadays, a lot of readily available information, however it needs processing. This is the case of the field of Quantitative Structure-Activity Relationships (QSAR), which exploits several thousand molecular descriptors, and quality control and multivariate calibration where hundreds of spectroscopic signals are easily obtained from spectroscopic methods. Genetic Algorithms, Simulated Annealing, and Tabu Search are some of the methods that are widely used to process available information to find sets of optimal models. In this case the problem that arises is how to compare the selected models. This work proposes a new measure of the distance between two models, and we will demonstrate that this model distance allows clusters of similar models to be found and the most diverse models to be caught in such a way as to preserve maximum information and diversity.

ORAL PRESENTATION

- **Pre-processing methods in Hasse Diagram methodology** (2004) *VI Partial Order Workshop* - Bayreuth, 15-16 April, 2004