Environmental Informatics, Environmetrics, Chemoinformatics, Chemometrics: Integration or Separation!?

Kristina Voigt
Helmholtz Zentrum Muenchen, German Research Center for Environmental Health (GmbH), Ingolstaedter Landstr. 1, 85764 Neuherberg, Germany,
e-mail: kvoigt@helmholtz-muenchen.de

Abstract: Several interdisciplinary scientific disciplines have been established in the past 30 years focusing on the computer science, mathematical and statistical aspects of environmental protection issues. An important society in this respect is the International Environmental Modelling and Software Society iEMSs who is organizing this conference. However, iEMSs will not be treated in this paper. Other disciplines concerning the evaluation of environmental and chemical data are: Environmental Informatics, Environmetrics as well as Chemoinformatics and Chemometrics. The aim of the research of these societies is to get more information out of the data and to arrive at some kind of decision support to improve the status of the environment. The introduced disciplines have all a different history, focus, grade of internationalisation and background of members. This paper aims at informing the members of the societies and interested working groups in the scientific vicinity as well as those generally interested in environmentally-related scientific disciplines.

Keywords: Environmental Informatics; Environmetrics; Chemometrics; Chemoinformatics; ranking methods, environmental databases, interdisciplinary environmental research

1. INTRODUCTION

The tremendous increase of environmental and chemical data, results in an increasing need for a new generation of software tools that assist the extraction of information from data and transforming the results into knowledge. In front of this background scientific disciplines evolved world-wide which are of great importance to solving the huge environmental data proliferation problems. The topics of Environmental Informatics, Environmetrics as well as Chemoinformatics and Chemometrics have been initiated in order to support the evaluation of these huge amounts of data. Around 40 years ago, both environmental science and computers were in the starting blocks of a world-changing evolution. In the status nascendi, these two fields seemed to be completely independent. Computer science with its technical and technological origin was and is opposed to the diverse scientific, ecological, social, and political environmental movements in many respects. Nevertheless, the increasing performance of computers and the need for solutions that respond to environmental stress let these different areas amalgamate. In the early till mid-1980s time was ripe for merging computer science with environmental science (Pillmann, 2006, 2007). The scientific discipline of Environmental Informatics was born. Concerning the data-analysis of environmental and chemical data the scientific field of environmental statistics also named Environmetrics was established in the late 1980s (El-Shaarawi, 2002). With respect to chemistry it was realized some decades ago
that the amount of information accumulated by chemists, in the long run, be made accessible to the scientific community only in electronic form, in other words, it had to be stored in databases. This new field, which deals with the storage, the manipulation, and the processing of chemical information, was emerging (Gasteiger, 2003). This is now one important field in Chemoinformatics. Since then, the application of computer science to the solution of chemical problems has ventured into many more areas of chemistry also into the field of environmental chemistry and ecological chemistry. The International Chemometrics Society was founded in 1974 and was rapidly developing in the 1980s (Einax, 2004).

Several different disciplines were emerging some twenty to thirty years ago which recognized the urgent need to combine aspects of computer science, statistics and mathematics with environmental sciences in general and environmental chemistry in particular. Although the need for these interdisciplinary and applied disciplines was recognized, collaborations between these groups are still in the initiating phase.

This article should support and trigger further understanding and hence cooperation among the disciplines Environmental Informatics, Environmetrics, Chemoinformatics and Chemometrics and of course trigger the interplay with the iEMSs.

2. ENVIRONMENTAL AND CHEMICAL DISCIPLINES WITH RESPECT TO COMPUTER SCIENCE AND STATISTICS

2.1 Introduction

As indicated above, the International Environmental Modelling and Software Society iEMSs who is organizing this conference is not treated explicitly in this paper but those societies and groups which are working in similar areas are discussed.

In the following section the environmental and chemical disciplines are introduced with their societies, WebPages, background literature as well as some journals in which the fields are covered. It goes without saying that only examples are given concerning the background literature and journals.

Table 1 gives an overview on the disciplines and societies.

<table>
<thead>
<tr>
<th>Name of Discipline</th>
<th>Society</th>
<th>URL</th>
<th>Background material Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemoinformatics</td>
<td>No society defined yet</td>
<td><a href="http://www.gdch.de/vas/tagungen/tag5316_e.htm">http://www.gdch.de/vas/tagungen/tag5316_e.htm</a></td>
<td>Handbook of Chemoinformatics (Gasteiger, 2003)</td>
</tr>
</tbody>
</table>

2.2 Environmental Informatics

Environmental problems in the 1980s resulted in an increased ecological awareness worldwide, which considerably enhanced political and scientific activities for the protection
of the environment. In the science sector, a rapidly growing community conceived new computer applications for decision making and information exchange in the field of environmental protection. Activities in the field of Computer Science applied to environmental research and protection are denoted by the term "Environmental Informatics" (EI). This is the global definition of EI. The EI community continually addresses new and challenging topics in the interface between the fields of Informatics (Computer Science) and Environmental Sciences. The nucleus of the upcoming field of "computer application for environmental protection" was the 1st Symposium held at the Karlsruhe Research Center in 1986 and the foundation of the Technical Committee "Informatik im Umweltschutz" (Environmental Informatics). The new extended title of the Technical Committee is now Environmental Informatics, Informatics for Environmental Protection, Sustainability and Risk Management. More than 22 years later, 22 EnviroInfo conferences have been held with approximately 3,000 papers by 7,500 authors published in proceedings’ volumes since the first event. From a German initiative the conference has largely expanded into the European field. The history and the development are outlined in Figure 1.

The scope of the Technical Committee in particular consists in the dissemination and application of modern methods and techniques of computer science in the environmental sector and to lend expert assistance in the development of effective, application specific informatics tools. The collaboration of partners from universities, research institutions, agencies and industries etc. is performed. Apart from the organization of conferences and proceedings’ volumes the Committee publishes a newsletter twice a year. In this publication news on the current work in the society, upcoming events like conferences, workshops are announced, conference reports are given. In the section scientific contributions new developments are explained. The newsletter which is written in German and partly in English is available on the Web page of the society (see Table 1). An important aspect of the work of Environmental Informatics is the establishment of special interest groups and working groups which take a closer and more in depth look at specific issues, e.g. Environmental Information Systems in Business, Simulation in Environmental and Geological Sciences, Environmental Databases, etc. These working groups meet regularly, publish their results in proceeding volumes, and report to the society. Apart from these basic activities, several textbooks have been written by members of the EI community over the past two decades. One of the first basic textbooks was entitled "Umweltinformatik" edited by Bernd Page and Lorenz Hilty in 1994. A second edition was published soon after the first one (Page and Hilty, 1995). An international textbook edited by Nicholas Avouris and Bernd Page followed soon after (Avouris and Page, 1995). Two
important books concerning one of the basic topics of EI, namely environmental databases and environmental information systems were published by Guenther (1998) and Rautenstrauch and Patig (2001), respectively. A recent attempt to extend the focus from environment to sustainable development in general was made by Hilty (2004). An approach to quantify the impact of information and communication technologies on sustainability was developed in a recent project (Hilty, 2006a). This is only a small sample of available background material concerning EI.

The Environmental Informatics initiative does not have a scientific journal of its own. However, a Special Issue entitled "Environmental Informatics" of the scientific journal Environmental Modelling and Software was published recently (Hilty, 2006b). The history, current research and trends in environmental informatics are given in this issue.

2.3 Environmetrics

TIES (The International Environmetrics Society) is a non-profit organization aimed to foster the development and use of statistical and other quantitative methods in the environmental sciences, environmental engineering and environmental monitoring and protection. To this end, the Society promotes the participation of statisticians, mathematicians, scientists and engineers in the solution of environmental problems and emphasizes the need for collaboration and for clear communication between individuals from different disciplines and between researchers and practitioners. The Society further promotes these objectives by conducting meetings and producing publications, and by encouraging a broad membership of statisticians, mathematicians, engineers, scientists and others interested in furthering the role of statistical and mathematical techniques in service to the environment. The first TIES Conference took place in Cairo, Egypt in 1989. TIES is an international society and its conferences are being held at several continents of the world. The 2007 conference was in Mikulov, Czech Republic, while the TIES 2008 conference will be held in Kelowna, BC, Canada.

Abdel El-Shaarawi the founder of the TIES conferences is optimistic about the future of the field of Environmetrics as quoted in an interview article (Esterby, 2003). Environmetrics has been around for a reasonably long time and has been quite successful expanding in terms of its methods and their range of application. Prof El-Shaarawi is of the opinion that without any doubt, statistics is the broadest field of science in terms of the use of its methods. The same formula for advancement as used by the great statisticians in the past, if used by the current generation, will guarantee a bright future for the field. This formula requires that statistical methods should be developed to address real problems in science and technology, and statisticians should be taught how to effectively communicate with scientific disciplines.

Currently there are ongoing activities that TIES will become a section of the International Statistical Institute (ISI).

An important background work of TIES is the Encyclopedia of Environmetrics, edited by Abed El-Shaarawi and Walter Piegorsch (El-Shaarawi, 2002). The publication of this major reference work, which includes more than 500 detailed articles, is a landmark for environmetrics. It really defines the subject and indicates its future directions. As can be seen, there is no shortage of real problems for scientists to work on, where their investigations have direct impact on improving the quality of life for the current and future generations.

The official journal of the International Environmetrics Society is named Environmetrics, published by Wiley-Interscience. The Editor-in-chief is Abdel El-Shaarawi (2008). In this journal selected papers from the TIES conferences are published.

TIES also publishes a newsletter which is available via the Webpage (see Table 1).

2.4 Chemoinformatics

Chemoinformatics is a fairly new name for a discipline that has been around for quite a while. Different people sometimes give rather different definitions of chemoinformatics.
The rather broad and general definition by Johann Gasteiger the editor of the Handbook of Chemoinformatics (Gasteiger, 2003) is: Chemoinformatics is the application of informatics methods to solve chemical problems. Chemoinformatics is the use of computer and informational techniques, applied to a range of problems in the field of chemistry. These in silico techniques are often used in pharmaceutical companies in the process of drug discovery.

The field of Chemoinformatics was not founded, nor was it formally installed. It slowly evolved from several, often quite humble, beginnings. Scientists in various fields of chemistry struggled to develop computer methods in order to manage the enormous amount of information and to find relationships between the structures and properties of a compound. In the 1960s some early developments became evident which led to a flurry of activities in the 1970s (Gasteiger, 2003). These days a big issue in this respect is the topic Chemoinformatics and Drug Development.

In Germany conferences on Chemoinformatics have been initiated and held by the CIC (Chemistry-Information-Computer - A division of the German Chemical Society). These conferences have international speakers as well as an international audience. Last year’s conference was the 3rd German Conference of Chemoinformatics / 21. CIC-Workshop. Important aspects of Chemoinformatics are published in the Handbook as well as in the Textbook of Chemoinformatics (Gasteiger, 2003). The field of chemometrics has seen considerable growth in 2006. This growth is evidenced by the fact that significant advances are no longer found in the pages of a few specialty publications but across a wide range of mainstream chemistry and general science journals, even in the Journal of the American Chemical Society JACS (Agrafiotis, 2007). So, no specific journal on chemoinformatics exists. However, in a variety of chemistry journals the aspects of chemoinformatics are published, e.g. Journal of Chemical Information and Modelling published by ACS (American Chemical Society).

2.5 Chemometrics

Chemometrics is the application of mathematical or statistical methods to chemical data. The International Chemometrics Society (ICS) offers the following definition: Chemometrics is the chemical discipline that uses mathematics, statistics, and formal logic:

- To design or select optimum experimental and measurement procedures
- To provide maximum relevant chemical information by analyzing chemical data
- To obtain knowledge about chemical systems.

Chemometrics, however, is a relatively young science that was born at the end of the 1960s and became an independent working field in the 1970s. The name "chemometrics" was first coined by the Swedish scientist Svante Wold in the early 1970s. His cooperation with the American analytical chemist Bruce R. Kowalski, who at the time was working on pattern recognition methods in chemistry, resulted in the foundation of the International Chemometrics Society in 1997. The first German working party on chemometrics was founded in 1984. Around that time chemometric working groups were also established in many other countries (Einax, 2004). The International Chemometrics Society has no active WebPage at the moment. In Germany the working group "Chemometrics and Laboratory Data Processing" within the division of Analytical Chemistry in the German Chemical Society (GDCh) emerged from the former working groups of Chemometrics and Laboratory Automation, which united in 1994. This group focuses on the combination of data acquisition in the laboratory, their statistical evaluation, plausibility checks and the appropriate quality assurance as good analysis practice (http://www.chemie.uni-jena.de/institute/ac/einax/ak_chemo/ak_chemo).

Current developments in chemometrics are characterized more by applications than by fundamental work in the statistical and/or mathematical field. In addition to classical problems of analytical chemistry, like the optimization methods, interdisciplinary works are gaining more and more interest. The investigation of the environment (Einax, 1997), of food and pharmaceuticals (Voigt, 2004, 2007), and chemometric modelling of technological processes and process analytical chemistry should be mentioned. A further important topic is the application of chemometrics to analytical quality assurance.
Concerning environmentally-directed chemometrics the reference work Chemometrics in Environmental Sciences (Einax, 1997) should be mentioned. A special issue of the journal Analytical and Bioanalytical Chemistry was recently issued (Einax, 2004) which describes the state of the art as well as current trends in chemometrics. Important journals in this field are: Chemometrics and Intelligent Laboratory Systems published by Elsevier, Journal of Chemometrics published by Wiley-Interscience, and Journal of Computational Chemistry published by Wiley-Interscience.

3. INTERACTION AND COOPERATION BETWEEN THE APPLIED ENVIRONMENTAL DISCIPLINES

All the discussed disciplines, as well as the organization hosting this conference, the International Environmental Modelling and Software Society iEMSs, are organized in international societies and working groups who work in theoretical fields and aim at making contributions to the better understanding of environmental problems and finally aiming at improving the environmental conditions of our planet. All groups are also aware and prepared to do some interdisciplinary work. Although there are a variety of overlapping activities, limited knowledge about the work of other societies and hence little cooperation between the societies has yet been performed. Very few attempts have been made so far to initiate common sessions at conferences or common workshops. At the EnvirolInfo 2003, 17th International Conference Informatics for Environmental Protection EnvirolInfo in Cottbus, Germany, a workshop on chemometrics was organized (Gnauck, 2003). At the EnvirolInfo in Vienna, Austria a TIES Session took place (Pillmann, 2002). These were single events and did not have the impact on the two societies and groups of experts as expected. So unfortunately these kind of cooperative activities have been discontinued so far. Of course quite a few experts attend several of the mentioned interdisciplinary conferences and present their results.

A new attempt is made at the iEMSS in Barcelona while organizing a session on Environmental Informatics, History and Current Trends by the author of this paper. Figure 2 should demonstrate the current status and the hopefully future trend in the co operations between the named groups and societies.

![Figure 2: Societies which Focus on Informatics, Mathematical and Statistical Aspects of Environmental Protection Issues.](image)

Currently the societies have more or less one direction: the improvement and protection of our environment by applying their theoretical knowledge. For the chemical societies this is only the case in those research areas where environmental topics, like environmental chemicals are treated. The aim could be reached more effectively if the cooperation between the groups were intensified.
It is envisaged that the societies and experts continue to expand their efforts in collaboration and taking advantages of other theoretical aspects and ideas concerning environmental protection.

4. APPLICATION EXAMPLE OF APPLYING ALL DISCIPLINES

This current research approach should demonstrate an example where environmental informatics, environmetrics on one side are combined with chemometrics and chemoinformatics on the other side. Furthermore, this example can also be regarded as an environmental modelling approach.

In this approach we evaluate 17 Internet databases (objects) by 16 pharmaceuticals (attributes/criteria). The method used for ranking databases is the Hasse Diagram Technique (HDT) which originates in discrete mathematics. It is a multi-criteria evaluation method which can well be used as a tool to rank objects, in this case databases, and is hence also applicable for decision making. The applied methodology reveals the best (maximal objects) and the worst (minimal objects) databases and conflicts among them, due to different information content.

![Figure 3: Hasse Diagram of 17 Databases x 16 Pharmaceuticals (equivalent object: \{CEX;RXL\})](image)

In Figure 3 we can explicitly see 16 out of the 17 discussed databases. The database CEX is equivalent to the database RXL. There exists only one maximal object, the equivalence class \{CEX;RXL\} which means that these two databases are better with respect to their availability on data of the 16 selected pharmaceuticals than all the other 15 databases. The minimal objects are ICS, OEK and INT. They are the worst databases in this evaluation approach (Voigt, 2007).

5. OUTLOOK

Interdisciplinary work, like given in the example which combines environmental informatics (databases) with environmetrics (mathematical method: Hasse Diagram Technique) and also lies in the chemometrics and chemoinformatics area (pharmaceuticals) helps to support the collaboration between the different societies named in this paper. Additionally the author is of the strong opinion that differences in the approaches and
scientific backgrounds should be used and cultivated in order to get a better understanding of the situation of our environment and hence being able to improve environmental conditions.

REFERENCES


Hilty, L.M., B. Page, J. Hrebicek (Guest Editors), Special Issue Environmental Informatics, *Environmental Modelling & Software*, 21 (11) 1517-1629, 2006b.


