Human and Ecological Risk Assessment, Expert Systems

Dušek L.1,2, Holoubek I.1, Jarkovský J.1,2, Koptíková J.2, Kohout J.1, Šmíd R.1, Pavliš P.2, Mužík J.2 and Hřebíček J.2

Abstract

The aim of this article is to introduce topic of conference tutorial focused on human and ecological risk assessment. Educational part of the tutorial is designed as methodical overview of risk assessment studies, exposure assessment and analyses of toxic biological effects. Data processing and associated information flow will be demonstrated in practical examples from environmental and epidemiological monitoring. All methodical aspects will be demonstrated and trained in two model case studies, prepared over representative and already published data. Both studies will be supported by information and expert systems that make data sources readily accessible for analyses. First case study is focused on environmental risk assessment and monitoring of surface water quality. Czech nationally distributed expert system TRITON® will be presented as a model solution for processing of very heterogeneous abiotic and biotic data. Seasonally repeated monitoring of physical environment and biological communities from more than 400 river profiles will be used to demonstrate advantages and limits of multivariate analog modelling (searching for and typology of reference sites, calibration of environmental gradient data, etc.). This case study will introduce educational block oriented on processing of biodiversity data. Second case study will address the problems of population human risk assessment and epidemiological studies. Cancer epidemiology was selected as a model system due to its societal importance. The study will be based on Czech national expert system (SVOD®) developed to process representative national cancer registry (standardized representative database collected since 1977) that allows quantitative evaluation of long-term epidemiological trends. Processing of population data using modern technologies (multivariate modelling and artificial intelligence techniques) will be demonstrated.

1. Introduction and aims of the tutorial

Ecological and human risk assessment can be characterized from the viewpoint of informatics as very complicated and non-standardized processing of heterogeneous data (mostly retrospectively collected from various sources) leading to probabilistic estimation of some uncertain (prospective approach) or on the other hand very certain (retrospective approach) risk event. In other words, there are many inputs required and only limited number of outputs provided, however always with strongly serious impact. Heterogeneity of input data ranging from controlled laboratory bio-tests to multilevel time- or space-aggregated data of course implies heterogeneity of applicable computational approaches. Computerised environmental information systems are rapidly being rolled out into use under such circumstances, but typically they are not complex enough to cover all data describing e.g. ecosystem level of organization. Lack of available software tools or algorithms mostly leads to simplified evaluation of some data components or to asymmetric assessment based more on easily accessible laboratory data. Evaluation of computational and expert capability of environmental information systems lags far behind their application and mostly we know very little about their impact on final estimation of risk.

This session is therefore not proposed only as blind presentation of different environmental information systems. The main ambition is to discuss the following problems that are commonly associated with or sometimes even generated by different types of information systems:

1 Research Centre for Environmental Chemistry and Ecotoxicology (RECETOX), Faculty of Science, Masaryk University in Brno, Czech Republic
2 Centre of Biostatistics and Analyses, Faculty of Medicine and Faculty of Science Masaryk University in Brno, Czech Republic
- how the information systems affect methodology of gathering of primary data
- the relation between costs and benefits in computer-assisted risk evaluation
- novel solutions for collection, aggregation and filtering of heterogeneous, large scale data (multi-dimensional systems, grid computational techniques, data mining, GIS methodology)
- user-oriented expert systems ensuring validated and flexible feedback of the assessor towards complexity and quality of primary data
- information systems and persistent organic pollutants (POPs) (data needs, modelling of persistence, multimedia models, expert services evaluating long-range transport of POPs, ...)
- implementation of the following important, but complicated fields in complex information systems (data models, computational and technological requirements):
  - processing, aggregation and evaluation of biotic data at population or ecosystem level
  - centralized procurement of large scaled data, retrospective analyses of environmental time series
  - processing of biodiversity data in environmental risk assessment
  - modelling suitable for risk assessment, generation and simulation of different scenarios (exposure pathways, toxic effects, ...)

Some of the barriers that limit complexity of environmental information systems are obvious. In contrast to relatively easy issue of novel bio-test or indication method, timescales are long and the ability to switch a system to more complex level is limited by cost and organisational constraints. In some key fields (biodiversity monitoring, fate of persistent compounds, population epidemiological risk) it is hard to collect representative data in sufficient period of time. Processing and interpretation of often incomplete data from natural systems cannot be straightforward and is dependent on effective computational techniques. It is mostly not possible to carry out adequate assessment of the large scale systems with techniques that have been successful in smaller systems with limited heterogeneity.

The tutorial is designed to provide general overview of standard methodology in risk assessment studies with subsequent practical training with simultaneous open discussion in the following fields:
- data management standards in context of risk assessment methodology
- analyses of spatial and temporal scale of hazardous situations
- development of risk assessment scenario and its execution from the viewpoint of risk assessment studies
- processing of biological data resulting from two major strategies in risk assessment: bio-tests performed on model systems and real monitoring at ecosystem level
- searching for reference standards in biomonitoring and epidemiological studies

We hope that this tutorial will help to reduce confusion and sometimes conflict over the roles of information technology in environmental and human risk assessment, and a lack of scientific as opposed to a managerial simplified approach in the development of expert functions. All the presentations will document that environmental informatics is a professional field, able to react on practical demands in even very complicated systems.

2. Model case studies in tutorial

All methodical aspects will be demonstrated and trained in two model case studies, prepared over representative and already published data. Both studies will be supported by information and expert systems that make data sources readily accessible for analyses and tests.
First case study is focused on environmental risk assessment and monitoring of surface water quality. Nationally distributed expert system TRITON® will be presented as a model solution for processing of very heterogeneous abiotic and biotic data. Seasonally repeated monitoring of physical environment and biological communities from more than 400 river profiles will be used to demonstrate advantages and limits of multivariate analog modelling (searching for and typology of reference sites, calibration of environmental gradient data, etc.). This case study will introduce educational block oriented on processing of biodiversity data.

Second case study will address the problems of population human risk assessment and epidemiological studies. Cancer epidemiology was selected as a model system due to its societal importance, recent development and also as a topic that is recently very dreaded by general public. The study will be based on Czech national expert system (SVOD®) developed to process representative national cancer registry (standardized representative database collected since 1977) that allows quantitative evaluation of long-term epidemiological trends and modelling of risk at population level. The study should demonstrate aggregation of data from various sources, their multivariate processing and final interpretation, that obviously can not be straightforward. Bioindication potential of population data depends on our ability to reasonably stratified subjects (mostly patients of endangered cohorts of people) according to “natural” risk factors, like regional attributes, life style or demographic structure.

The tutorial introduces ecological and human risk assessment methodology in logic consecutive steps, from auditing of data sources through exposure assessment and effect testing to final risk estimates. Each of the steps is to be questioned from the viewpoint of required data inputs, their processing and outputs. Lectures and demonstrations will be focused namely on practical aspects associated with information background of the process: practical data gathering and handling, “case-report forms” and data models, data analysis in common tools and finally, the interpretation of end-points. This methodical outline will be reflected in complex typology of different risk situations (hazard identification) and clearly preset scenario for further testing and assessment (risk assessment in experimental design). In order to fulfill methodical standards, each case study is arranged in five major blocks, each related to special methodical segment:
- block 1: hazard identification, conceptual model and risk assessment scenario
- block 2: exposure assessment, exposure pathways, quantity and time profile of exposure
- block 3: evaluation of biological effects
- block 4: risk characterization
- block 5: methodical outputs, scenario and design for further evaluation (monitoring)

Practical demonstrations should accelerate acceptance of methodical standards namely in the following fields:
- Data models, auditing and validation of information sources. Relevant arrangement of input information, description of the situation, hazard identification.
- Regionally specific aggregation of accessible data, quality control and GIS model of the area of interest, mapping of exposure pathways and levels.
- Hierarchical structure and prioritization of biological indicators according to ecological criteria, susceptibility to stress factors and accessibility for measurements. Application of data from large-scale biomonitoring networks. Searching for biological “hot spots” and reference standards for biological systems in the area of interest. Selection of proper biological receptors and endpoints.
- Role of laboratory (eco-toxicological) tests, critical overview of commonly used parameters from these tests (NOEC, LOEC, NOAEL, LOAEL, IDx, LDx, EDx, ICx, LCx, ECx). Implementation in practical decision making.

950
Benchmarking of regionally available chemical and biological data, risk characterization and regionally specific interpretation.

3. Informatics for ecological and human risk assessment

Informatics is incorporated in each methodical step of formal risk assessment, from data gathering and analyses, through risk estimation to final validation of results and minimization of uncertainties. Information technologies assist in building of assessment scenario and in optimization of experimental design. Finally, we can not forget communication of the results. To summarize, information technologies and data analysis give the assessment process all its desirable properties:

- evidence-based background
- sufficient information power
- credibility and confidence
- effectiveness and presentation skills

Evidence-based methodology. Based from the viewpoint of informatics, environmental risk assessment can be characterized as methodically diverse processing of heterogeneous data (mostly retrospectively gathered from numerous sources) leading to probabilistic characterization of some uncertain (mostly prospectively recognized) risk event or effect. In other words, there are numerous inputs required and only limited number of outputs provided in the “computational black box” of risk assessment. Heterogeneity of input data ranges from controlled laboratory bio-tests to multilevel ecosystem descriptors measured repeatedly both in time and space. Although heterogeneity of inputs could imply heterogeneity of applicable computational approaches, one of the substantial roles of informatics is to standardize the aggregation and subsequent analyses of data. We should accept data management rules as indispensable part of risk assessment methodology that brings “evidence – based” approach and results.

Standardized data management. “Evidence-based” methodology is built on prospectively planned sampling and experimental design, with statistical reasoning of sample size and objective targeting in the area or landscape of interest. Description of risk situation should be based on representatively described environmental components and all potentially stress factors. Parametric structure of endpoints is preset as representative for important biological receptors and absolute measures are evaluated with respect to objective benchmarks. Recent informatics and statistics operates with technologies and algorithms that are able to standardize multi-level process like environmental risk assessment. The requirement to control all steps that rely on heterogeneous data surely results in increasing pressure on development of robust data management systems. Development of standardized and automated tools for data management is of increasing importance also due to strong impact on decision making about key sources of environmental pollution.

Indispensable computational methodology and data processing. The principal role of informatics in environmental risk assessment cannot be narrowed only to standardized gathering and aggregation of primary data. All phases of the process (hazard identification, exposure assessment, dose-response monitoring) are intrinsically associated with some level of uncertainty and so the final conclusions are based on stochastic analytic methods. At this point we must accentuate key role of GIS technology, multivariate processing of ecological bioindicators and finally very important dose-response modeling and probabilistic characterization of risk. Each of these methods, and many others as well, represent unique field of computational science with its own methodology and progress. Notwithstanding the methodical variety, ecological risk assessment must assimilate only verified approaches with sufficiently robust algorithms, suitable to heterogeneous or incomplete data. Reliable computational methodology bridges the gap between environmental and experimental data and makes the whole process as effective as it is possible.
Flexibility to incorporate new parameters and technologies. Similarly like other biological sciences, recent development in ecotoxicology is attacked by occurring high-throughput technologies that accelerate toxicity testing and push forward our understanding of mechanisms of eco-toxicity. The expected seamy side of methodical development is overproduction of experimental data, very often measured in artificial conditions and under influence of environmentally irrelevant doses of stress factors. The role of informatics is to incorporate this novel experimental data in the frame of routine environmental assessment. The applications however cannot be blind and must stand on proper dose-response measurements treated with respect to inherent uncertainty (extrapolation from artificial conditions, inter-species variability, concentration levels, ...). As an example we could mention choices that are relatively very easy when dealing with high-dose exposures but that become substantially more difficult in low-dose range of exposures. Another example may be the advantage we could take from fascinating development in the field of genomics and budding toxicogenomics that can contribute to environmental risk assessment of carcinogenic compounds. Computational sciences indeed can open the door for completely new parameters, approaches or testing technologies even without violation of standard routines.

Information systems. Very important contribution of informatics to environmental risk assessment that can never be neglected is data and information service delivered directly to the point of decision through information systems. The systems substantially enhance individual susceptibility into risk assessment and ensure rational comparative evaluation or benchmarking in networks of differently experienced users. Only accessible information can form the basis for decision making and only sophisticated information systems can guarantee accessibility of information from large (bio)monitoring programmes, etc.. Informatics viewed in terms of information systems also brings important training and educational platform.

Expert systems. The last aspect that should be mentioned in the context of environmental informatics is expert information safety, availability and accessibility. Hundreds of environmental monitoring projects had been performed (and currently are carried out) in tens of countries and indeed it seems that there is even redundancy of environmental data in European region. The critical problem however is the availability of correct interpretation of data. Outputs of governmental and non-governmental projects typically exist in separated databases of many institutions and their merged browsing and analyses are complicated and require participation of specialists from many fields and scientific knowledge of respective data as well. This time consuming process often strongly limits the availability of reasonable data at the point of strategic need. Namely data on environmental fate of persistent organic pollutants requires long-term analyses and are difficult to interpret without multiple entries. Therefore, there is strong demand on expert information systems. The question is, what are the barriers for routine processing of already estimated data? And what is the safe and relevant information service for general public and how to ensure it? All these problems can be technically solved through safe expert systems that aggregate relevant data sources with final automated analytic tools. This is the future of responsible “environmental reporting”.

The study and issue of all educational materials was supported by research project INCHEMBIOL, Ministry of Education Czech Republic project no. 002162412 (RECETOX, Masaryk University, Brno). Human risk assessment study and expert system SVOD are granted by Ministry of Health Care, Czech Republic and by project MZO 00209805 solved in Masaryk Memorial Cancer Institute, Brno.