Research Centre for Environmental Chemistry and Ecotoxicology
Faculty of Science, Masaryk University
Kamenice 126/3, 625 00 Brno, CZ; www.recetox.muni.cz

EcoRA
Methodology of Ecological Risk Assessment - Tutorial

Methodology, Tools, Applications and Case Studies

Edited by Ladislav Dušek, Jiří Jarkovský and Ivan Holoubek
June 2007, Brno, Czech Republic
Methodology of Ecological Risk Assessment - Tutorial

Authors:
Ladislav Dušek
Jiří Jarkovský
Ivan Holoubek
Vratislav Kubík
Jiří Hřebiček
Jaroslav Ráček
Pavel Čupr
Klára Hilscherová
Eva Gelnarová
Danka Němethová
Jakub Hofman
Jana Klánová
Jana Koptíková
Michal Hejč

Editors:
Ladislav Dušek
Jiří Jarkovský
Ivan Holoubek

Design:
Vratislav Kubík

Research centre for Environmental Chemistry and Ecotoxicology,
Faculty of Science, Masaryk University.
The main scientific, research and educative activities of RECETOX are focused on environmental chemistry, ecotoxicology and risk assessment with special attention to persistent, bioaccumulative and toxic substances (PBTs).
The basic scientific topic is the study of relationships between environmental occurrence and levels of various types of pollutants.

Institute of Biostatistics and Analyses,
Faculty of Medicine and Faculty of Science, Masaryk University
is a research department oriented to solution of scientific projects and related services, especially in the field of environmental and clinical data analysis, organisation and management of clinical trials, software development and ICT applications.
The activities of IBA are focused mainly on organisational and expert services for large scientific projects in environmental or clinical research.

June 2007, Brno, Czech Republic

This report should be cited as follows:
METHODOLOGY OF ECOLOGICAL RISK ASSESSMENT:
TUTORIAL

CONTENTS

Ecological risk assessment: Theoretical background ................................................................. 4
Introduction .................................................................................................................................. 4
Basic definition of Ecological Risk Assessment ....................................................................... 4
Main information sources for EcoRA methodology used in this tutorial................................. 7
Theoretical background: Environmental and Human Risk Assessment and the Contribution of ICTs 8
Abstract ...................................................................................................................................... 8
Risk assessment studies and their methodology .................................................................... 8
Tiered analyses in environmental risk assessment studies .................................................... 9
Information and communication technologies in environmental and human risk assessment .... 10
More pragmatic view into the complicated EcoRA methodology ......................................... 13
Most important aspects of analysis of environmental data and related research topics .......... 17
Development of environmental information systems and related research topics .................. 20
Conclusion .................................................................................................................................. 20

Theoretical background: Current Trends in Environmental Modelling with Uncertainties ... 21
Abstract ...................................................................................................................................... 21
Introduction .................................................................................................................................. 21
Uncertainty Analysis with Using Computer Algebra Based Systems ......................................... 22
Uncertainty Analysis Support with Using CAS Maple Tools .................................................. 23
Case Study: Air Pollution by the Transport in the Czech Republic ........................................ 24
Case Study: Assessment of Waste Management Indicators in South Moravia Region .......... 25
Conclusions .................................................................................................................................. 26
References .................................................................................................................................... 27

I. Problem formulation ............................................................................................................... 29
“Zlin” Case Study .................................................................................................................... 29

Problem formulation of “Zlin” case study ............................................................................... 29
Definition and goals of the study ............................................................................................ 29
Informational background ....................................................................................................... 29
Selected stressors ..................................................................................................................... 30

1
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Studied region and its description</td>
<td>31</td>
</tr>
<tr>
<td>Contamination sources</td>
<td>33</td>
</tr>
<tr>
<td>Biological receptors</td>
<td>35</td>
</tr>
<tr>
<td>Conclusion</td>
<td>37</td>
</tr>
<tr>
<td>Enclosure: A short Guide to GIS for EcoRA users</td>
<td>39</td>
</tr>
<tr>
<td>II. Exposure assessment</td>
<td>40</td>
</tr>
<tr>
<td>Design of sampling network</td>
<td>40</td>
</tr>
<tr>
<td>Sampling and laboratory procedures</td>
<td>41</td>
</tr>
<tr>
<td>Key pollutants: survey of environmental concentration levels and the flooding influence</td>
<td>43</td>
</tr>
<tr>
<td>Key pollutants: Local hotspots and Zlín urban agglomeration</td>
<td>49</td>
</tr>
<tr>
<td>Conclusions</td>
<td>52</td>
</tr>
<tr>
<td>III. Effect Assessment</td>
<td>53</td>
</tr>
<tr>
<td>Relevant receptors</td>
<td>53</td>
</tr>
<tr>
<td>Biomonitoring network: reference based approach for effect assessment</td>
<td>55</td>
</tr>
<tr>
<td>Data sources</td>
<td>55</td>
</tr>
<tr>
<td>Biological communities as assessment endpoint in the EcoRA</td>
<td>55</td>
</tr>
<tr>
<td>Statistical methodology</td>
<td>56</td>
</tr>
<tr>
<td>Approaches in analysis of biodiversity data</td>
<td>56</td>
</tr>
<tr>
<td>Biodiversity sampling</td>
<td>57</td>
</tr>
<tr>
<td>Diversity indices</td>
<td>58</td>
</tr>
<tr>
<td>Variability and testing of diversity indices</td>
<td>60</td>
</tr>
<tr>
<td>Species abundance models</td>
<td>60</td>
</tr>
<tr>
<td>Multimetric comparison with reference condition</td>
<td>61</td>
</tr>
<tr>
<td>Multivariate probabilistic comparison with reference status</td>
<td>62</td>
</tr>
<tr>
<td>Results</td>
<td>66</td>
</tr>
<tr>
<td>Conclusion</td>
<td>67</td>
</tr>
<tr>
<td>Acknowledgment</td>
<td>67</td>
</tr>
<tr>
<td>References</td>
<td>67</td>
</tr>
<tr>
<td>Ecotoxicological tests: evidence-based approach</td>
<td>69</td>
</tr>
<tr>
<td>Introduction</td>
<td>69</td>
</tr>
<tr>
<td>Processing of dose response data</td>
<td>69</td>
</tr>
<tr>
<td>Benchmark dose calculation</td>
<td>70</td>
</tr>
<tr>
<td>Software ADRIS</td>
<td>71</td>
</tr>
</tbody>
</table>
ECOLOGICAL RISK ASSESSMENT: THEORETICAL BACKGROUND

INTRODUCTION

The ecological risk assessment is a rather complex process, resulting finally into the assessment of stress factor influence on ecosystems or on their components. At present, these methods represent the most suitable and an increasingly employed tool which is used both in the Czech Republic and abroad to assess an existing or potential risk and the negative impact of anthropogenic and non-anthropogenic activities on the ecosystem’s structure and function.

This tutorial is primarily aimed to provide essential information on ecological risk assessment to students marginally concerned with this issue, as well as to the professional and lay public.

This text provides a basic introduction of methodology and gives an overview of tools facilitating the risk assessment. A case study is also provided, giving an example of applying individual stages of EcoRA procedure on real data.

This tutorial has been created at the RECETOX research centre, cooperating with the Institute of Biostatistics and Analyses of MU’s Faculty of Medicine and Faculty of Science.

The main information source for this tutorial was “Methodology of Ecological Risk Assessment” published in 2002 as a final report of the project IDRIS II (VAV 340/1/00) by TOCOEN, s.r.o., which is a complete methodology for EcoRA process together with the core publications in this field.

This tutorial is supported by the project of the Ministry of Education of the Czech Republic (project No. MSM 0021622412 INCHEMBIOL)

BASIC DEFINITION OF ECOLOGICAL RISK ASSESSMENT

The Ecological Risk Assessment (EcoRA) provides tools to assess negative impacts of natural disasters, human activities and pollutants on organisms or on entire ecosystems.

The EcoRA helps us to acquire, establish, comprehend and process the available information and to make it relevant for responsible decisions in environmental issues. These decisions may include, for example, measures to reduce the production of pollutants, regulations in the application of specific substances or recommendations to remedy a specific situation.

The risk assessment has been accepted as a fundamental tool for decisions in environmental issues thanks to incontestable advantages of this process:

- It provides quantified outputs to compare risks and to determine priorities. Actually, the risk cannot be entirely eliminated in most cases; when making the decision, it is important to find the most suitable solution from several different points of view.

- It represents a systematic tool for a better comprehension of risks. The uncertainty analysis in individual stages facilitates the identification of issues which require more attention.

- The risk assessment takes into account the uncertainties in the prediction of future environmental conditions, thus making the outputs more plausible.

- The risk assessment makes a clear distinction between the scientific part, which assesses the magnitude or probability of effects (risk assessment) from the other part, which considers the
proposed solutions and the risk acceptability (risk management). Potential errors are therefore limited and the reliability is enhanced.

- Although EcoRA has been primarily designed to assess the influence of chemical substances, this process can be successfully applied on a number of other stressors (floods, fires, climate change).

The EcoRA’s objective is to ensure the protection of environment, being based on a scientific assessment of ecological impacts of human activity or other influencing factors. This task can be only successfully accomplished if we define exactly the area of interest and suggest methods to determine the conditions and to quantify the expected changes. We must also be able to recognize the moment when those changes are significant and to assess correctly the relevance of uncertainty in our data and models.

From the point of view of defining exact and suitable objectives, endpoints, receptors, exposure assessment and effects assessment, the EcoRA is rather more complex than the human risk assessment. We need to answer questions such as: should we put more stress on endpoints on the “ecosystem” level, such as the biodiversity or production? Should we limit the study on species highly appreciated from the aesthetic and/or commercial point of view? Should we focus on representative species? Once we work it out and determine “what” to assess, another problem will arise soon: we need to choose the most suitable method “how” to do it.

The conception of ecological risk assessment is based on four essential parts (stages). This structure is more or less followed in all available methodical procedures. The basic conception thus involves problem formulation, exposure assessment, effects assessment, and risk characterization.

The problem formulation involves the identification of objectives and assessed endpoints, the preparation of conceptual model and the establishment of analyses plan. In this stage, the core of the problem is characterized; potential stressors are described, as well as their sources, the region of interest and potential risks. A rich informational background must be established in this stage, in order to facilitate the decision process in the subsequent stages, and to make our work easier then.
The **exposure assessment** is one of two analytical stages of the process. The aim of this stage is to make the best possible description of ways how the monitored parameters can be exposed to specific stressors and how to quantify this exposure. The input data can be, for example, real environmental concentrations in a specific matrix, or data from contaminants determination in organisms’ tissues. Various exposition models can be very useful.

The **effects assessment** is another analytical stage. Its objective is to characterize the relationship between the exposure to the stressor and the stressor’s influence on the assessed organism. The effects assessment might be based on epidemiological studies; in most cases, however, the information from toxicity testing is used. The values obtained in this stage will be employed in the subsequent risk assessment.

The **risk characterization** is a summary output of risk assessment, resulting from the data integration from previous stages. The effects of stressor exposure – with the objective to describe the risk – can be defined as the surface of a four-dimensional space. The individual dimensions are:

1. concentration of stressor to which the receptor is exposed,
2. time of exposure,
3. receptor’s properties,
4. form of effect.

In addition, the risk characterization involves the uncertainty assessment as well as the consideration of available evidences before giving a final output.

**Risk description** provides important information to interpret the risk results and identifies a threshold for adverse effects on the assessment endpoints. When evaluating ecological risks and the potential for response alternatives to achieve acceptable levels of protection, risk managers should characterize risk in terms of magnitude, severity, distribution, and the possibility of recovery for the affected receptors.

**Risk management**: The ecological risk assessment should find out whether a risk is present and to define a magnitude of that risk. With this information, risk managers have to integrate the risk assessment results with other considerations to make and justify risk management decisions. Other considerations in making risk management decisions include existing background levels of contamination, available cleanup technologies, and costs of alternative actions and remedy selections.

**Risk communication**: It is the process of informing people about the hazards of monitored site. The purpose of this step is to help residents of affected communities to:

- understand risk assessment and risk management,
- form scientifically valid perceptions of the likely hazards,
- participate in making decisions about how risk should be managed.
The process of informing people about site hazards is a two-way conversation in which the site team informs and is informed by affected community members.

**Main Information Sources for EcoRA Methodology Used in This Tutorial**


THEORETICAL BACKGROUND: ENVIRONMENTAL AND HUMAN RISK ASSESSMENT AND THE CONTRIBUTION OF ICTs

Ladislav Dušek, Ivan Holoubek, Jana Koptíková, Jiří Jarkovský, Vratislav Kubík, Jaroslav Ráček, Jiří Hřebiček

ABSTRACT

This chapter is aimed to introduce the topic of human and environmental risk assessment from the viewpoint of information a communication technologies and to provide basic overview of consecutive methodical steps in this multitask process. Simplified definition of most important methodical components is explained with special focus on exposure assessment and the analyses of toxic biological effects. Data processing and optimized information management flow are indispensable parts of all the steps and bear responsibility not only for common outcomes like data summaries and risk characterization. Information and communication technologies take important role from the very beginning of the assessment process (experimental design, sampling plans, and scenarios) to the last point including an analysis of uncertainties. Information and communication technologies and data analyses form a base namely for the empirical phases of the process and determine the success of environmental monitoring, multivariate modelling and reliability of the estimates of final measures. In the case of biological effect evaluation, the attention is paid namely to the dose response modelling while environmental monitoring and exposure assessment work predominantly with multivariate factorial or discrimination models. Environmental risk assessment can theoretically employ any type of biological data from molecular markers to biodiversity indices and therefore, none method of data analysis can be excluded from the usable list. That is why we rather tried to provide general methodical scheme than detailed listing of all available analytical techniques.

RISK ASSESSMENT STUDIES AND THEIR METHODOLOGY

Environmental and human risk assessment can be characterized from the viewpoint of informatics as complicated processing of heterogeneous data (mostly retrospectively collected from various sources) leading to probabilistic estimation of some uncertain (prospective approach) or on the other hand relatively certain (retrospective approach) risk event. Key methodical steps of the whole process are summarized in Fig. 1 and can be simply defined as follows:

![Environmental risk assessment (EcoRA) and key methodical steps](image-url)
1. **Problem formulation and hazard identification** - Introduction to any reasonably designed study. It includes recognition of the area of interest, collection and aggregation of required information and preliminary focus on identified principal pollutants (stressors), source of contamination and most vulnerable environmental components and biological receptors.

2. **Multi-component exposure assessment** - exploration, identification and quantification of important exposure pathways. It includes modelling and summaries of accessible data as well as empirical estimates of environmental concentrations of proposed key pollutants.

3. **Biological effect evaluation** - Empirical phase focused on concentration-related or dose-related reactions of biological systems. Principal aim is to get parametric measures that identify biologically dangerous concentration levels. The process should not be limited only to laboratory testing, it works with ecosystem monitoring as well. Whenever we have access to regional or national bio-monitoring network we should use this data as very powerful information background.

4. **Risk characterization** - is completely computational process that leads to the probabilistic estimate of the risk. In fact it is stochastic aggregation of data from all the preceding methodical blocks.

In other words, there are many inputs required and only limited number of outputs provided, however always with serious impact. The whole process can be also visualized as a circle that takes very different data and aggregates them according to given rules (Fig. 2).

---

**Figure 2** EcoRA methodology visualized as circle

The circle is constrained to rotate due to never-ending chain of problems entering the system. The image of „circle process“ is functional because not all phases must be necessarily employed in any type of problems. Different situations give to different phases very different weights. For example, the situation is thoroughly different if someone needs retrospective or prospective assessment than if it is sudden catastrophic situation like industrial accident or flood where we must in first line mitigate the immediate effects.

**TIERED ANALYSES IN ENVIRONMENTAL RISK ASSESSMENT STUDIES**

Heterogeneity of input data ranging from controlled laboratory bio-tests to multilevel time- or space-aggregated data implies heterogeneity of applicable experimental approaches. It is often very difficult to select the best approach and many later steps assume some specific result or outcome from the preceding
measurements. That is why the experimental plans must be substantially more flexible as we can compare with some other types of research, for example with clinical trials. We must solve the complexity of the system, where no unique, definitely the best model can be recommended. We can of course move on according to plan, but from the first step we must categorize the problem - and here rather multiple (independent) lines of evidence are better than any single approach. This simply minimizes the probability of missing or omitting of some important facts and makes the solution sufficiently robust.

We adopt simple methodology from the first step and re-organize it in so called tiers as it is depicted in the scheme of Fig. 3. The consecutive tiers are gradually unravelled according to methodology, and repeated if the uncertainty is too high. As an example, we can perform some screening biological test to survey and point out the sites of interest, but then – in testing according to given scenario - similar tests can be repeated to verify the findings or to prove it in the other sites. Tiered style allows us to control the process and to react flexibly on unexpected changes or situations.

Figure 3) Tiered style of EcoRA methodology

INFORMATION AND COMMUNICATION TECHNOLOGIES IN ENVIRONMENTAL AND HUMAN RISK ASSESSMENT

Information and communication technologies (ICTs) are 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.

Fig. 4 brings most important methodical streams with localization in EcoRA information flow. ICTs assist in building of assessment scenario and are indispensable in optimizing of experimental design. 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.
Some important methodical tools and gains are summarized in the following paragraphs:

- **Evidence-based methodology.** The “computational black box” of risk assessment covers a wide 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 must accept data management rules as inherent part of the methodology that brings “evidence – based” approach and results.

- **Standardized data management.** The evidence-based methodology is based on prospectively planned sampling and experimental design, with statistical reasoning of sample size and with objective targeting in the area of interest. The description of risk situation should be based on representatively described environmental components and all potentially influential stress factors. The parametric structure of endpoints is preset as representative for important biological receptors and absolute measures are evaluated with respect to objective benchmarks. Current informatics and statistics operate with technologies and algorithms that are able to standardize multi-level process like environmental risk assessment. Development of standardized and automated tools for data management is of increasing importance also due to strong impact on decision making process that follows the risk assessment.

- **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 environmental bioindicators and finally very important dose-response modelling and probabilistic characterization of risk. Each of these methods represent unique field of computational science with its own background. Notwithstanding the methodical variety, environmental risk assessment must assimilate only verified approaches with sufficiently robust algorithms, suitable even 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, current 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 seamy side of the development is overproduction of experimental data that are typically measured in artificial conditions and under influence of environmentally irrelevant doses of stress factors. The role of informatics is to incorporate this novel experimental dimension 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 science indeed is the only change how to open the door for these completely new parameters without violation of standard routines.

- **Development of information systems.** Very important contribution of informatics to environmental risk assessment that can never be neglected when data is to be automatically delivered directly to the point of decision. 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 programs. 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 service. 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 at least in our European region. The critical problem however is availability and correct interpretation of the 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. 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. 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”.

Writing about ICTs in the EcoRA methodology, we should never forget communication of the results. Risk estimate itself is complicated end-point, conditioned by many items and circumstances. Risk should be communicated as probability of some hazard event that however can reach different levels for different biological systems. Without adequate presentation, there is a danger of incorrect interpretation and the examined situation can be unacceptably underestimated or omitted. ICTs have the power to emphasise the right findings to the right people. We can make the results accessible even for general public or to provide interactive tools for the evaluation of different scenarios. There are numerous target groups for the
communication of EcoRA results with very different requirements on the form and content of EcoRA reporting (Fig. 5). Only modern communication technologies including open public web portals can serve for such complicated communication field.

![Diagram](image)

Figure 5) Communication demands on information service from EcoRA

**MORE PRAGMATIC VIEW INTO THE COMPLICATED ECORA METHODOLOGY**

Although the methodology seems to be relatively complicated and rigid, it is not true. It really has to be easily adapted for the assessment of many types of ecosystems like soil, surface fresh waters or some special type of urban area. In practical realization, each methodical block is responsible for answering of some principal questions and provides estimates with minimized uncertainties and risk of bias. Following paragraphs summarizes such simplified, question-oriented approach.

Any risk assessment starts with problem definition in several prominent components. First, we must know the situation where we are and at least approximately what has happened and what is required. Second, we must, at least hypothetically, identify potential source or problem. Once we know these two components we should concentrate on the following key questions that are to be answered, as based on the assignment:

- Are we able to prioritize further methodical steps?
- Are we able to define sampling and measurement plan?
- Are we able to stratify the area of interest according to risk?
- Are we able to recognize risk situation?

These questions target the scenario of further assessment. The answers of course require sufficient awareness of the situation in the area of interest that can be get from accessible databases and libraries or from pilot screening survey (screening of exposure pathways, pilot investigation of potential sources of toxic compounds, etc.).

That is why the process of assessment is unavoidably initiated by rather inventive searching for information where we typically work with three types of data sources (Fig. 6):

- widely accessible databases of toxic compounds and their characteristics,
- local databases and standards,
- pilot results of some tests and ad hoc measurements.
Figure 6) Problem definition and applicable data sources

Gained information must be processed and the sources must be prioritized, filtered and then aggregated in order to get really usable input into assessment scenario. The information processing is typically projected into the area of interest and so most recommended tool for this phase is some type of geographical information system (GIS).

The assessor must be aware of relative spatial and temporal scales of the hazardous situation being assessed. The segmentation of the area of interest determines the extent and range of exposure and identifies the sites with increased probability of risk attack. However, we must take into consideration the assignment for the assessment itself, it means to correlate between measurement feasibility and risk situation and assessment interest that is placed into by experts, public, environmental activists, owners or managers as it is indicated in Fig. 7.

Figure 7) Problem definition generates comprehensive situation plan

The situation plan thus opens following important questions that can be answered only in comprehensive experimental plan:
• How to target heterogeneous area of interest? And what is optimal and safe sampling plan for in situ investigation?
• What is the optimal set of further analyses?
• How to effectively combine ecological and chemical measures and laboratory tests?

The assessment scenario can be regarded as milestone of the assessment process, because it definitely summarizes obtained information and generates hypotheses and strict plan of further movement. Fig. 8 depicts basic principle of categorization of the scenarios working with primary and secondary exposure pathways. The idea of this approach is simple – most of the further measurements are determined by the type of exposed environmental matrices. If we are able to distinguish contaminated matrices, it simply places the necessary applied tests. The approach also implies primary and secondary risk for prospective studies. Distinguished separation of different types of ecosystems (matrices) can also be easily associated with legislation and guidelines valid for the field.

Figure 8) Assessment scenario and basic principle: „Where is the problem”

Here we must strongly emphasize very important dimension of any practically targeted risk assessment scenario, it means effectiveness and pragmatic simplicity. EcoRa is very laborious work and so there is only limited space for research activities. We cannot rely on scientific and progressive data trying or even hoping that it improves our understanding common ecological data. Instead of that, the simplest and already verified set of methods should be applied in first line and additional analyses should be applied only in the case of remarkable uncertainties.

The selection of proper biological end-points is the last important step that should be prescribed in the scenario. This step is extremely important because the risk estimate cannot rely only on chemical data:

• biological tests can reveal some new effects that cannot be predicted from chemical data (this holds namely for heterogeneous environmental mixtures of chemicals)
• bioindication or monitoring survey can indicate the effects in past, that cannot be detected by chemical analyses
There are basically two strategies in biological effects assessment (Fig. 9): (1) model testing under defined conditions (mostly laboratory tests and bioassays) and (2) monitoring of real ecosystems that is often referred to as bioindication. Typically we employ different tests or tests using different biological models (receptors) to cover the heterogeneity and various food strategies of biological systems in the field.

In final risk characterization chemical and biological estimates are evaluated by the help of probabilistic models:

\[ HQ = \frac{AEC}{TEC} \]

\[ \text{Hazard quotient} = \frac{\text{Actual environmental concentration}}{\text{Toxicological effect concentration}} \]

Figure 9) Two basic strategies for the selection of biological end-point

Figure 10) Simplified example that demonstrates risk estimate

Multiple biological tests of course require a wide spectrum of analytical methods as it will be explained in next paragraphs. At this point we should only emphasize, that outcomes of both chemical measurements (mostly applied during area recognition and exposure assessment – see Fig. 1, 2) and biological tests provide principally the same numerical output, concentration measure:
environmental concentration (mostly obtained on the basis of chemical monitoring) as estimate of the real (or predicted) contamination of target matrices

biologically safe concentration (obtained from laboratory tests, biomonitoring and/or in situ assays) as the estimate of the concentration that does not cause biologically harmful effect or on the other hand, the concentration that causes some degree of the effect.

Both types of concentration estimates are then mutually related in probabilistic models or ratios to estimate the probability of risk event (Fig. 10). We typically mutually relate two or more probabilistic distributional models and try to find concentration cut-off point that determines the highest acceptable environmental concentration that is still biologically safe.

MOST IMPORTANT ASPECTS OF ANALYSIS OF ENVIRONMENTAL DATA AND RELATED RESEARCH TOPICS

Data analysis plays strategic role 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 models 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. Currently, more and more sophisticated techniques are being applied in this field, namely data mining technology.

![Diagram](image)

Figure 11) Standard output of bio-tests: dose-response curve

- Standardization of laboratory tests and their outputs (i.e. concentration measures that are related to some level of biological effects: NOEC, LOEC, NOAEL, LOAEL, IDx, LDx, EDx, ICx, LCx, ECx). Practical implementation of these outputs in decision making includes following topics (see also examples in Fig. 11-13):
  - standardized and reliably estimated concentration measures
  - estimation and formal mathematical description of different types of dose-response curves (because the shape of the relationship is important as well)
- typology of dose-response curves for different situations and stressors
- mathematical processing of complicated, but real dose-response patterns (see Fig. 13).

Figure 12) Dose-response analyses and analytic solution for the data

- Benchmarking of ecosystem abiotic and biotic characteristics, risk characterization and regionally specific interpretation. This approach generates more complicated data structure than laboratory biotests with the following specifics:
  - It is impractical to measure these complex parameters in predictive design, where the conclusion must be made before any large-scale release and effects can occur, the evaluation is then retrospective not prospective.
  - In fact it is virtual reconstruction of the conditions of the injured environment prior to the pollution release so that it can be compared to the injured condition.
  - In the case of such complex parameters, standard statistical methodology fails or we have no time to verify hypotheses in repeated experiments.
  - An alternative methodology to common statistical scoring must be used and it is multivariate statistical modelling: a) processing of so called calibration data set that is representative for the ecosystem or b) work with environmental gradient data. We rank hazardous and clean sites using a linear discrimination with the support of training set of well characterized sites that are assigned to categories on a scale hazard. The site attributes are used to establish a discriminant function. If the function separates the categories sufficiently, new sites are classified on the hazard scale by applying the discriminant function (see Fig. 14, 15).
Figure 13. „Real” variability of dose-response curves and related parameters

Figure 14. Scheme of model processing of ecosystem data

Figure 15. Modes of multivariate processing of ecosystem data
DEVELOPMENT OF ENVIRONMENTAL INFORMATION SYSTEMS AND RELATED RESEARCH TOPICS

Computerized environmental information systems are rapidly being rolled out into use, but typically they are not complex enough to cover all attributes describing ecosystem level of organization. Lack of available software tools or algorithms mostly leads to simplified evaluation of some 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. The main ambition of informatics in this field is then to solve the following problems that are commonly associated with or sometimes even generated by different types of information systems:

- how the information systems affect methodology of gathering of primary data
- how to effectively optimize collection, aggregation and filtering of heterogeneous, large scale data (multidimensional systems, grid computational techniques, data mining, GIS methodology)
- how to develop user-oriented expert systems ensuring validated and flexible feedback of the assessor towards complexity and quality of primary data
- how to optimize information systems for monitoring of persistent organic pollutants (POPs) (data needs, modelling of persistence, multimedia models, expert services evaluating long-range transport of POPs, specific biological effects, etc.)
- how to maximize ration between costs and benefits in computer-assisted risk evaluation
- how to implement information systems in the following important, but complicated fields (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
  - 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 organizational constraints. In some key fields (biodiversity monitoring, fate of persistent compounds, population epidemiological risk) it is hard to collect representative data in relatively short 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.

CONCLUSION

We hope that this short overview 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 systems. Many of the mentioned problems are in detail characterized in the next chapters of the proceedings.
THEORETICAL BACKGROUND: CURRENT TRENDS IN ENVIRONMENTAL MODELLING WITH UNCERTAINTIES

Jiří Hřebiček, Michal Hejč

ABSTRACT

The characteristic feature of current environmental modelling is the complexity and uncertainty of its mathematical representation (uncertainty of formula). Imprecision of its input data is another characteristic feature, where it is not possible to omit influences of primary monitoring (e.g. gaps of data, errors of measuring facilities, human factor, etc). Nowadays, information and communication technology (ICT) capabilities are growing rapidly and applied mathematical software (e.g. computer algebra systems, statistical packages, etc) becomes more powerful to overcome problems with formula complexity and uncertainty. The basic methods how to deal with the data uncertainties are well known and standardized from the last century, but some of their comparisons and recommendations for environmental modelling are not known enough. Paper presents generalized approach and shows universal methodology how to use computer algebra based systems (CAS) for the implementation of mathematical models with formula and data uncertainties. Various approaches for solving uncertainty with the computer algebra system Maple are presented. The modified model COPERT III with included uncertainty was developed in Maple and some results of the case study for the transport influence and all the related air pollution in the Czech Republic are presented together with the case study devoted to the assessment of waste management indicators in South Moravia region.

INTRODUCTION

Environmental modelling with uncertainty has to consider three interrelated categories [5], [24]: model framework uncertainty, data uncertainty and application niche uncertainty.

1. **Model framework uncertainty**: means uncertainty in the underlying science, the system of governing equations that make up the mathematical model and developed algorithms of this model which are the result of incomplete scientific data or lack of knowledge about the factors that control the behaviour of the system being modelled. This can also be the result of simplifications necessary to translate the conceptual model into mathematical terms.

2. **Data uncertainty**: is caused by measurement errors, analytical imprecision and limited sample sizes during the monitoring or collection and treatment of data. Over the years, many different approaches to evaluating and expressing the uncertainty of measurement results have been used. The end result of international collaboration with the various national metrology institutes working in ISO/TAG 4/WG 3 was Guide to the Expression of Uncertainty in Measurement (or GUM as it is now often called). Data uncertainty is sometimes referred to as reducible uncertainty because it can be minimized with further study [23].

3. **Application niche uncertainty**: means uncertainty regarding the appropriate application of a model, e.g. using certain ICT tools. This is therefore a function of the appropriateness of a model for use under a specific set of conditions, [22]. It is particularly important when choosing among existing models for an application outside of the problem for which it was originally developed and/or developing a larger/better model from several existing models with different spatial or temporal scales.

In the paper we introduce a new generalized approach to environmental modelling with uncertainty with using computer algebra based systems (CAS). The approach is applicable in all fields of environmental modelling (waste management, air quality, water quality, etc.) and even in some non-environmental areas, where some huge amounts of data with uncertainties are involved (medicine, economy, etc.).
Further we present some results of the case study of the comparison of various approaches solving uncertainty with the CAS Maple to get better results for specific problem of all transport influence on air pollution in the Czech Republic.

Other case study about solving waste management data uncertainties has been published [10]. The study uses the same modular principle, but since the amount of the processed data is not as enormous, it does not make use of CAS, so the paper will not discuss this case study in the same depth as previous. General principle of environmental modelling with uncertainties is shown here instead to illustrate its use in this case study.

Uncertainty Analysis with Using Computer Algebra Based Systems

Computer algebra based systems involve the direct symbolic and algebraic computation (SAC) of the governing equations of mathematical models of environmental problem and also the estimation of the sensitivity and uncertainty of model outputs with respect to model inputs. The symbolic technology allows CAS to maintain all of the essential mathematical knowledge and structure inherent in a formula, equation, model, or program. Consequently, SAC can apply rules of mathematics to environmental problems and quickly produce answers that are much more meaningful than just numbers or graphs. For example, often, the actual solution of the problem is not the final step as many applications require further mathematical processing after a given solution computation. SAC approach provides greater flexibility for post-processing because the system maintains the history of its computation and is not just a black box. CAS maintains all of the underlying mathematical structure including those of previous expressions that were used to create an expression. By applying the associated packages of mathematical and graphical operations, one can analyze sensitivities of parameters, convergence of solutions, parametric dependencies, and much more.

Today there are various CAS, from the utilities for just one application area to complex systems, which can perform computations from all areas of mathematics [6]. The known SAC systems are e.g. Yacas, HartMath, The OpenXM project, Prologie, GiNaC, ArtLandia, Axiom, CoCoA, Derive, Algebra Domain Constructor, Fermat, GAP, GANITH, GRG, GRTensor, LiDIA, GNU DOE Maxima, Magma, Maple, Mathematica, Mathomatic, MathSoft, MATLAB, MathTensor, Milo, MP, MuPAD, NTL, Pari, Reduce, Schur, Singular, SymbMath, TI-92 Calculator, and TI-92 Plus.

Environmental modelling with uncertainty with using CAS issues usually from following approaches:

1. **Interval arithmetic**: is used to address data uncertainty that arises either due to imprecise measurements or due to the existence of several alternative methods, techniques of theories to estimate parameters [17]. The primary advantage of interval mathematics is that it can address problems of uncertainty analysis that cannot be studied through probabilistic analysis. It may be useful for cases in which the probability distributions of the inputs are not known.

2. **Fuzzy theory**: is a method that facilitates uncertainty analysis of systems where uncertainty arises due to vagueness or fuzziness rather than due to randomness alone, (http://www.uncertainty-in-engineering.net/). It is used for treating uncertainty of the parameters which are constant, not random, but impossible to measure and its value is not mathematical form. Further applications of fuzzy randomness on environmental problems may be found in [18], etc.

3. **Probabilistic analysis**: is the most widely used method for characterizing uncertainty in environmental models, especially when estimates of the probability distributions of uncertain parameters are available [8]. This approach can describe uncertainty arising from stochastic disturbances, variability conditions, and risk considerations. In this approach, the uncertainties associated with the model inputs are described by probability distributions, and the objective is to estimate the output probability distributions.
4. **Methodology of Checkland or Soft Systems Methodology (SSM):** is iterative approach consists of seven distinct stages, forming a life cycle of environmental model with uncertainty [4]. In the last few years, the SSM was currently used for describing the river ecosystem in India [2].

The traditional role of the ICT in environmental modelling has been the solution of mathematical models [7]. Even some of the most sophisticated and expensive modelling systems are essentially numerical solvers more elaborate user interfaces and distributed calculation via Internet.

The process of environmental modelling using CAS consists of the spiral cycle (IDENTIFY – DEVELOP – IMPLEMENT – SOLVE – ANALYZE – MODIFY), see Fig. 1, which shows the way how complex CAS automate all phases of environmental modelling, [13].

![Diagram of environmental modelling using CAS](image)

**Figure 1) Life cycle (spiral) of environmental modelling using SAC**

Further, we will concentrate on using Maple (http://www.maplesoft.com), which we have used since 1991, (http://www.solvingproblems.ethz.ch). It has own programming language and export its worksheets into HTML, MathML, LaTeX, RTF, and XML format files or C#, Fortran, Java and MS Visual Basic languages. Its suitable tools for network communication enable connecting Maple to processes on remote hosts on a network (such as an Intranet or the Internet) and exchange data or program codes with these processes.

**Uncertainty Analysis Support with Using CAS Maple Tools**

Appropriate Maple tools dealing with the uncertainty are following (http://www.maplesoft.com):

1. **Interval arithmetic:** The package intpakX of Maple provides basic data types and operations for interval arithmetic as well as additional features for further interval computation [21]. It contains the type checking functions, all arithmetic functions including powers, trigonometric and hyperbolic ones, set operations on the interval, range operations for a given function, complex
number support and some basic numeric methods as the Newton’s method for finding a root of an uncertain function.

2. **Fuzzy theory**: The Fuzzy Sets toolbox of Maple allows constructing and working with fuzzy subsets of both the real line and of user-defined finite sets. Its modules automatically generate fuzzy controllers from a collection of user-defined rules. This allows modelling, testing, and modifying fuzzy systems in the interactive Maple worksheet environment.

3. **Probabilistic analysis**: The ScientificErrorAnalysis package of Maple provides representation and construction of numerical quantities in Maple that has a central value and associated uncertainty or error, which is some measure of the degree of precision to which the quantity’s value is known. The associated uncertainty can be specified in absolute, relative, or units in the least digit form. In the returned object, the uncertainty is quantified in absolute form. This package allows various first-order statistical calculations of uncertainty analysis with these quantities and methods following the GUM guide.

4. **Getting the online data and program codes**: The Sockets package of Maple allows getting data and program codes for the computation online from the web [6]. In particular, it enables two independent Maple processes running on different computers on a network to communicate with one another.

Of course, the Checkland’s SSM is not automatically implemented in Maple yet.

**CASE STUDY: AIR POLLUTION BY THE TRANSPORT IN THE CZECH REPUBLIC**

The emissions from transport in the Czech Republic has been analyzed with respect to uncertainties using ICT tools of Maple [13], [21], where the implemented mathematical model of transport air emissions in Maple has issued from the well-known mathematical model COPERT III [19]. Therefore, we will not describe its mathematical equations here. The COPERT III methodology is assumed to reflect real world conditions, but it is not fully clear from its documentation to what extent fuel consumption estimates have been based on official test cycle results and to what extent they are based on measurement of real world cycles. Currently, the EU funded projects [1] and [20] are further extending the knowledge on emission factors for all transport modes and all pollutants and we have taken into account their results and recommendations. We have introduced another set of emission factors for the computations at the local level. These factors do not represent the pollutant emission from one kilogram of fuel, but from travelling one kilometre. It looks more suitable to make comparison of these two emission factor sets, but the final answer is still opened. The data for emission inventory have been used from the Czech Hydrometeorological Institute and the Centre of Transport Research, which have competence to provide emission inventory for the Czech Republic.

The original COPERT III methodology [19] has been improved with respect to possibilities of uncertainty analysis in Maple, for the calculation of emissions, which are measured and statistically estimated. Uncertainty in emission factors have been considered, together with total fuel consumption for relevant transport modes, which markedly improves the calculation accuracy of emissions produced by transport. Selected emission factors, which are based on measured values, used probabilistic approach, and further the direct dependence of the relationship of transport performances given in passenger kilometres or ton-kilometres were eliminated.

The chosen results presented on Fig. 2 (CO2 and NOx, emissions in kg/inhabitant) were obtained from Maple program after two iterations of the SSM.
Figure 2) CO₂ and NOₓ emissions (kg/inhabitant) generated in the Czech Republic by all types of transport (individual and public passenger, road goods, urban public, rail diesel traction, waterway and air).

CASE STUDY: ASSESSMENT OF WASTE MANAGEMENT INDICATORS IN SOUTH MORAVIA REGION

Waste management data have been collected from waste generators and treatment facilities in South Moravia Region to its new district state administrative offices and the regional government annually since 2000. There are used similar information systems as in the Slovak Republic described by [9] and [16]. Waste management indicators are assessed from these data.

This case study presents the use of the modular approach, which divides the annually collected waste management data into particular modules described by [10]. Every module processes one portion of the data (portions may have intersections) and uses one method for uncertainty processing (interval arithmetic, probability approach, fuzzy approach or their combination).

The dependencies between these modules are illustrated in the Fig. 3 and described further with examples from the case study [10].

Figure 3) Modules’ dependencies

Basic modules are formed by models of some portions of the waste management data. It’s possible to say, that the model is a function of primary collected waste management data and the information from knowledge database. The model also contains some parameters, which are the subjects of optimization after evaluation. The value, which is an output of this function, can substitute some values from primary waste data.

The model of municipal waste is an appropriate example to explain this methodology. The model uses the number of inhabitants dNI and evaluates the coefficients of production per capita dp1 in the district d from its district primary collected data, uses archived coefficients of production per capita dp2 and year average
production dp3 in the district d from the last year, etc. The output of this model is estimated the annual regional municipal waste production
\[
\sum_d \bar{d} N_d p_d
\]

and the annual regional municipal waste production from the last year:
\[
\sum_d \bar{d} N_d^2 p_d, \sum_d \bar{d} p_d
\]

The module for Trustworthiness of the data estimates trustworthiness of primary collected data (it can be of the type fuzzy, interval or probability), and it takes into consideration primary data from past three years, models, knowledge and some parameters, which are suitable for optimization.

The probabilistic approach is often used and data are clustered with respect to all subjects (villages/cities) with wrong data in the last year, all data differing from interval models (i.e. minimum and maximum of municipals waste in given subject), etc. Particular probabilities are multiplied and cluster parameters are subjects of optimization. This is not a mandatory rule; probabilities can be substituted by fuzzy approach, multiplication by other arithmetic, etc.

The module Result just computes the result from primary waste management data, its trustworthiness and the model. Result takes all their optimization parameters and the output of this function is a value which will substitute the origin value in waste database, but with the confirmation of the government response officers in the case of sensitive data (e.g. demolitions waste, hazardous waste, etc.). Substituted values are mostly the same as in the primary data (more the data are uncertain means more substituted values). This function applied on all values produces the new database.

This new database is evaluated. During evaluation some evaluation criteria (again it can be used with some parameters) are used if possible. Evaluated database is checked by the check function (which is a function of the Result and some knowledge). In accordance with evaluation criteria the optimum database is founded in sufficient number of iterations, generally in cooperation with decision makers from the government of South Moravia.

Only thing which now remains to clear is the schedule of optimization. In this case study the optimization has been done manually by experts in waste management field, which gave the process some reasonable sets of parameters and the best of them (in terms of optimization process) have been used.

Next step will be the use of some sophisticated procedures to schedule optimization process. CAS are one of options, others being SSM, genetic algorithms, etc.

**Conclusions**

Our several years’ funded research at Masaryk University [11], [12], [13], [21] has shown that introducing uncertainty into environmental modelling is suitable. Deeper knowledge of the mathematical model and the data together with uncertainty and sensitivity analysis can show how much the input uncertainty influence the outcome of the model. Classification of the parameters and the data into subsets (where some of them are sufficient to be known roughly and some of them more accurately) can divide the problem of uncertainty into parts, solved by different approaches (interval arithmetic, fuzzy and probabilistic theory).

Moreover, for our case study we did not find any parameter at the modified model COPERT III where the uncertainty damages the output results totally [13]. This means all of input model parameters can be handled as uncertain with more or less accuracy. Some results do not match the reality because of numerous simplifications. Using the Checkland’s SSM we were able to find out which changes are necessary. In our case
study, this shown us that only some of the simplifications we made are not desirable and if we make the next iteration, the outcome will be more similar to the original model than this version.

In the second case study we discussed the process of assessment of waste management indicators of South Moravia region, where we has been collected primary waste management data since 2002. The new approach which takes into account uncertainty of data can generate better results. But the exploration is far to be complete at this moment and there is still much to do – especially in the field of particular waste modules.

There is still a lot of further research in the area of environmental modelling with uncertainty, but we hope this presented investigation helped a little at least to give the future research the proper direction - to have environmental models with acceptable complexity, covering the influence factors and using current ICT.

REFERENCES

1. ARTEMIS (http://www.trl.co.uk/artemis/)


20. PARTICULATES (http://vergina.eng.auth.gr/mec/lat/particulates/)


I. PROBLEM FORMULATION

“Zlín” Case Study

The tutorial follows the EcoRA methodology on case study of Zlín Region where several types of EcoRA scenarios were employed for the solution of environmental questions.

All scenarios are documented in EcoRA steps: starting with problem formulation, exposure and effect assessment, and ending with risk characterization. These scenarios are presented using a set of software from the databases and statistical packages to GIS analysis and visualization.

Problem Formulation of “Zlín” Case Study

Definition and Goals of the Study

The Zlín Region is an important region of the Czech Republic with high economical and cultural significance and more than 200 000 inhabitants. The results of human industrial and other activities are significant sources of environmental pollution which can affect the ecological status of the region. Moreover, the area is exposed to periodical floods which can significantly affect the distribution and availability of pollutants. The study should answer the following questions:

- The importance of floods for ecological risks
- The importance of local hot-spots
- The importance of urban agglomeration

Informational Background

The extensive evaluation of available information is crucial in the first phase of EcoRA problem formulation. Both information from database and from screening examination of area of interest can be applied. Many
databases and information sources were used in this study. Data from a number of different institutions was obtained - especially data about our area of interest.

The data for exposure assessment were gained from our monitoring. As for the effect assessment, data from different available databases were used.

**MAIN EXTERNAL SOURCES:**

- **CENIA (CZECH ENVIRONMENTAL INFORMATION AGENCY)**
  - basic provider of GIS layers used in this study. This institution stores spatial data not only about basic characteristic of the region, but about many specific and detailed characteristics of the region, too.

- **AOPK (AGENCY FOR NATURE CONSERVATION AND LANDSCAPE PROTECTION OF THE CR)**
  - The main scope of activities involves mapping various species in the Czech Republic, monitoring the biodiversity and landscape, and protecting natural reserves.
  - The data from AOPK about soil properties were used in this study.

- **ČHMÚ (CZECH HYDROMETEOROLOGICAL INSTITUTE)**
  - ČHMÚ is concerned with the monitoring of the emissions (imissions) and atmospheric deposition, air conditions, climate, water quality, hydrological balance, pollution sources...
  - To some extent, we used climatological data about precipitation and temperatures, as well as hydrological data about flow rate.

- **ZLÍN REGIONAL AUTHORITY**
  - Monitoring of our sampling sites was performed in cooperation with Zlín Regional Authority.
  - Data about pollution sources as well as information about regional environment status were obtained from this institution, too.

**SELECTED STRESSORS**

Polycyclic aromatic hydrocarbons (PAHs) and polychlorinated biphenyls (PCBs) and organochlorinated pesticides (OCPs) were selected as the primary stressors. Many different sources of these substances are present in our study region and high concentrations of these pollutants were measured in floodwater, too.

PAHs, PCBs and OCPs together with PCDDs and PCDFs form a wide group of POPs (Persistent Organic Pollutants). All these substances are characteristic by many common properties which make them potentially dangerous to the environment. POPs are stable, non-degradable substances, which can be transported over long distances. They are notorious for their capability of bioaccumulation and biomagnification. Due to their high lipophilicity and low solubility in water, they bind to the soil (sediment) organic matter. Therefore, sediments or soils can serve as a sink for these hydrophobic organic contaminants and can be considered as sources of secondary pollution.

The danger of POPs lies in a number of harmful effects for organisms (toxic, mutagenic and carcinogenic effects, endocrine disruptors, effects on reproduction and foetal development).

PAHs and PCBs OCPs were chosen because of

- occurrence of hot-spot sources by the industrial establishments or in the rural area,
- more extensive contamination in Zlín – Otrokovice agglomeration,
• floods, which can significantly influence the contamination in the flooded area, due to properties of these pollutants.

The main sources of the POPs contamination in the Zlin Region are industry (chemical, plastics-and-rubber, boot-and-shoe, engineering), fuel combustion, transportation, agriculture, dumping grounds and waste waters.

Heavy metals were selected as a second group of monitored stressors. These elements and their compounds are found naturally in the environment; however, anthropogenic activities increase their environmental concentrations to hazardous levels. Heavy metals are persistent and non-degradable, have a high degree of bioaccumulation potential and have toxic effects on organisms.

With regard to increased environmental concentrations, Anthracene, Benzo-a-pyrene, Fluoranthen (PAHs) and DDT (OCPs) were chosen from POPs for an example risk characterization in our study.

From heavy metals, representatives of the most dangerous metals for the environment (Lead and Cadmium) were selected.

**STUDIED REGION AND ITS DESCRIPTION**

The simplest way of description of a studied region is the application of available GIS layers. Integration of the spatial and temporal information from these different layers provides a detailed characterization of the study area. Our example project presents many descriptive layers of the Zlin Region which make you acquainted with its characteristics. The map below provides information on the basic structure of this region (inhabited/uninhabited area, streams, transport infrastructure).

![Map of the Zlin Region](image)

*Figure 1) The layers “Zlin_region”, “Urban_areas”, “Roads”, “Rail_road”, “Rivers”, “Water_areas” and “Forests” are displayed in this map. It gives us an overview about the basic structure of the Zlin Region.*

Displaying the layers “Population” or “Protected areas” will help us to estimate which parts of the protected areas or how many inhabitants are potentially most endangered (especially in relation to the sources).
Figure 2) The layers “Cont. sources”, “Rivers” and “Population density” are visualized in this case. We have obtained the map representing the relation between the location of pollution sources and the number of potentially endangered inhabitants. This is a very important piece of information, especially for human risk assessment.

Zlin Region is located in the south-eastern part of the Czech Republic, which is located in Central Europe. This region with 10 cities and 72 villages has about 197 000 inhabitants, and a density of 190 inhabitants per km². The largest city is Zlin (83 000 inhabitants). This area has high economic and cultural significance and is significant for its industrial and agricultural activities. A great deal of the Zlin Region is formed by highlands, especially on the East. This hilly landscape plays a key role in the formation of local climate and transportation of pollutants.

The altitude is 183 m above sea level in the lowest point (by the Morava River) and 835 m above sea level at its highest point.

The whole area pertains to the Outer Carpathians geological unit. The basement rock is flysch (Paleogen age) – a complex of the sedimentary rocks (claystone and sandstone) with a thickness of more than 500 m.

From the pedological point of view, this region is considerably heterogenic (Czernozems, haplic Luvisols, rendzic Leptosols, Albeluvisols, Cambisols). Our sample sites are formed from fluvisols. These are specific alluvial soils developed in river flood plains.

The main geomorphological units are Lower Moravian Vale, Vizovice Highlands, Mladcov Highlands, Hostyn Hills and Slovak-Moravian Carpathians. The study area belong to the basin of Danube, the most considerable river is Morava with its tributary Drevnice.

The landscape is formed predominantly from woodland (42.2 %) and farmland (43.6 %). The region is poor on mineral sources; only brick-clay, gravel and stone deposits are significant. On the contrast, the area is rich in mineral waters sources.
Three basic climate zones are characteristic for this region. The western part of the region – around Morava river – forms the Zone A, a warm area with mild winters. The major part of the region belongs to the mildly warm Zone B. The zone C (moderately cold area) is located in higher altitudes. The long-term annual mean air temperature varies from 7 to 8 °C. The long-term annual precipitation in higher altitudes is about 900 mm, in lower positions it ranges from 625 to 800 mm. The air humidity ranges from 72 to 77%; the air flow is south-western and north-western.

In terms of phytogeography, this region belongs to the Mesophyticum carpatium unit and there are supracoline and submontane vegetation belts.

The Zlin Region had been formerly considered as a stable industrial zone; after the privatisation in the 1990s, this position was shattered. A great deal of economically active population is employed in industry (34.8%) and services, only 2.7% in agriculture. The main industry branches are chemical, plastics-and-rubber, boot-and-shoe, food processing industry and engineering. Depending on altitude, the main agricultural crops are potatoes, beetroot, wheat, barley, maize and fodder crops. Livestock breeding is typical for higher altitudes.

This study is focused on the periodically flooded area around the rivers Morava and Drevnice. Since 1996, monitoring has been conducted as repeated sampling and measuring of fluvisols and sediments. The sampling sites were assorted into five subregions with different characteristics. This area presents a suitable model ecosystem for pollutant accumulation, migration, transformation and distribution and dislocation research at a local and regional level.

In July 1997, the region was affected by disastrous floods. These caused great material and ecological damages and took several lives. Enormous rainfall plagued the whole Odra River basin, the north part of the Morava River basin and partially the Elbe River basin (15 000 km²). The total estimated precipitation volume in this five-day period was 3 milliards m³, i.e. 1/20 of total annual precipitation volume in the Czech Republic and 1/5 of the outflow of water volume from The Czech Republic. In addition, the situation on Morava River was complicated by a second flood wave, 10 days later. In historical context, this was a totally rare event but due to incontinently human landscape interventions, similar events are expected to be more frequent.

The majority of our sample sites were flooded (often for more than several days), long-time sediments were washed up and silt materials were deposited up to several centimetres deep. Because of earlier monitoring of this area, this situation brought an opportunity to evaluate the presence of contaminants in sample sites after the floods and compare results.

**Contamination sources**

The area is endangered by numerous potential pollution sources; there is over 50 potential or verified sources of industrial pollution, about 10 potential sources of agricultural pollution and over 40 million m³ of sewage per year are drained to the surface waters without treatment. Moreover, there are about 40 controlled, illegal or old dumping places, which present a risk to underground and surface waters.

In the flooded areas, high concentrations were measured of PCBs, PAHs and other POPs and heavy metals in soils, sediments and surface water. In most cases, these substances originated from industrial facilities, agriculture, solid fuels combustion and transport. Floods in 1997 increased the influence of these sources: many of them were flooded and washed up. The most dangerous sources of pollution were: Hamag (foundry), Barum Continental (rubber industry) Moravan (machine building, automobile equipments), Toma (energy industry) and Zlinská doprava (transportation). Water treatment plants were disabled and flooded, too, which contributed even more to the contamination.
Figure 3) Map of the main sources of the pollutions. The companies Barum Continental and Hamag were found as the most important sources of pollution.
Figure 4) Orthophoto map of the chosen hot-spots (Barum Continental, Hamag) with sampling sites indication. The closeness of the rivers increases the possibility of the contaminants washing out by the floods.

**BIOLOGICAL RECEPTORS**

The right selection of biological receptors has to follow strict rules to fulfil the EcoRA problem definition. Many prerequisites of receptor need to be fulfilled, as is shown on the following picture.
Figure 5) Required characteristics of receptors

Another consideration on the receptor selection is its level in the ecosystem due to various impacts of stressors on different ecosystem levels.

Figure 5) Ecosystem levels for receptor selection
The main aim of this study was to evaluate the ecological condition of the alluvial soils with regard to sediments (as the secondary sources), primary sources and the floods as the key factor in the mobility of the pollutants.

Therefore, receptors important for soil processes and soil quality were chosen. The effects will be assessed on the organisms, communities and ecosystem level and the representatives of producers (Poaceae), destruents (soil microorganisms) and consumers (Lumbricidae, Enchytraeidae and Isostomidae).

SOIL MICROORGANISMS perform many essential activities in the soil – organic matter decomposition, new substances production, soil structure creation... They have a considerable influence on nutrient flow and energy circulation in the soil. Moreover, the microbial biomass is a significant source of the nutrients for the other organisms in the terrestrial environment. The condition of the soil microbial communities tells us a lot about the soil quality and fertility.

THE FAMILY LUMBRICIDAE was chosen as a representative of the earthworms. They play a significant role in the soil food web, are important producers of the soil organic material and their activities influence soil properties and structure. The advantage of earthworms is their frequent usage in soil ecotoxicology, thus many standardized tests and toxicological data are available.

**CONCLUSION**

Our study area represents a suitable model for monitoring of the contaminated environment. It is a region with high industrial and agricultural activities, thus different sources of contaminants are present and we can study pollutants accumulation, migration, distribution or distribution at a local and regional level. On the other hand, the areas with background levels of the contamination occur in our study region too.
Many industrial facilities are situated in the agglomeration Zlín – Otrokovice. The most frequent output pollutants from these companies are PAHs, PCBs, AOX and heavy metals.

On the one hand we can study the influence of the particular sources (hot – spots), on the other hand we can observe the effects of the contamination on a regional level.

Additional important factor for our case study are the floods. Because of the monitoring of the area before and after the flood, it brought us the opportunity to evaluate the impact of this disaster on the distribution and effects of the pollutants (see figure below).
ENCLOSURE: A SHORT GUIDE TO GIS FOR ECO_RA USERS

Geographic Information System (GIS) is a specialized, database-driven computer information system. The database contains information on spatially distributed features, activities or events, which can be defined in space as points, lines or areas. The GIS allows users to capture, store, display, manipulate and analyze geographically linked data. Risk decision-makers can use GIS to discover and demonstrate spatial relationships, making GIS a valuable tool to assess the risk of monitored stressors.

In the ecological risk assessment, GIS can serve as a powerful tool to display and analyze data during the planning, scoping, and problem formulation stages, during the exposure assessment, and displaying and evaluating the results of the risk characterization. It is also a very helpful tool to communicate the information to risk managers and other stakeholders.

Once the risk assessors decide to use a GIS, they need to choose the software. A number of GIS software is available (commercial or open source). A key role for choosing GIS software is to consider the required functions. Important functional capabilities to consider include: data capture, data storage, data management, data retrieval, data analysis, and data display. Software ESRI – ArcMap 9.0 was used in this study. This text describes the basic possibilities of GIS usage in EcoRA process. Work with the attached GIS file will be presented, too.

Example GIS files “EcoRA” (mxd, pmf) together with the geodatabase file (mdb) are attached. You can browse the “pmf” file (Published Map Documents) in ArcReader – it is freely available to download at: http://www.esri.com/software/arcgis/arcreader/download.html. The “mxd” file (Arc Map Documents) can be opened in ArcMap 9.x. It provides you with many tools to view, modify, analyze and model geographic data. Of course, you can use other GIS software too, all the presented data is saved in the “mdb” file for these purposes.
II. EXPOSURE ASSESSMENT

In many cases, the results in this chapter are based on the published paper “Hilscherová, K., Dušek, L., Kubík, V., Hofman, J., Cupr, P., Klanova, J., Holoubek I.: Redistribution of Organic Pollutants in River Sediments and Alluvial Soils Related to Major Floods. Journal of soils and sediments 7 (3), 167 - 177, 2007” with some supplementary subsequent analyses.

DESIGN OF SAMPLING NETWORK

The preliminary screening research and database information aggregation revealed 5 sub-regions with different characteristics of soil and sediments and pattern of contamination. The set of 14 localities (Figure 1, 2, Table 1) for both soils and sediments was established in the region and sampled regularly throughout 1996-2005.

Figure 1) Location of the soil and sediment sampling sites and the study regions (labeled I. – V.)
Figure 2) Location of the sampling sites with regard to the main point-sources of the contamination.

<table>
<thead>
<tr>
<th>N.</th>
<th>Name</th>
<th>Associated sites</th>
<th>Reasoning</th>
<th>Potential source of contamination</th>
<th>Additional characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Braterejovka &amp; Lutoninka upper stream</td>
<td>Sed1, S1, S2, S3</td>
<td>Local background</td>
<td>A few small sources (SFC)</td>
<td>P 1650, A 26 km², AFR &lt;0.5 m³/s</td>
</tr>
<tr>
<td>II</td>
<td>Lutoninka &amp; Drevnice upper stream</td>
<td>Sed2, Sed3, Sed4 Sed5, S4, S5</td>
<td>Local contamination</td>
<td>AG, chem., industry, SFC, OEO, SW, TF</td>
<td>P 10500, A 70 km², AFR 0.6 m³/s</td>
</tr>
<tr>
<td>III</td>
<td>Drevnice</td>
<td>Sed6, Sed7, Sed8, Sed9, S6, S7, S8</td>
<td>Industrial centre contamination</td>
<td>AG, chem., energy and other industry, metals manuf., OEO, SFC, SW, TF</td>
<td>P 93000, A 183 km², AFR 2.2 m³/s</td>
</tr>
<tr>
<td>IV</td>
<td>Morava - upper stream</td>
<td>Sed10, Sed11, S9, S10, S11, S12</td>
<td>Regional background</td>
<td>AG, chem., and other industry, OEO, SW</td>
<td>P 15000, A 35 km², AFR = 54 m³/s</td>
</tr>
<tr>
<td>V</td>
<td>Morava - industrial agglomeration</td>
<td>Sed12, Sed13, Sed14, S13, S14</td>
<td>Industrial centre contamination</td>
<td>AG, chem., energy, food and other industry, SW, SFC, TF, OEO</td>
<td>P 46000, A 100 km², AFR = 57 m³/s</td>
</tr>
</tbody>
</table>

AG = agriculture, OEO = old ecological onuses, SFC = solid fuels combustion, SW = sewage waters, TF = traffic emissions
A = area, AFR = average flow rate, P = population

Table 1) Description of selected localities

**SAMPLING AND LABORATORY PROCEDURES**

Surface sediments (from top 10 cm layer) were collected using a trowel from the sedimentation basis of the riverbed in zones of calm flow, close to the river bank. Representative samples were prepared by mixing 5-8
subsamples from area about 4 m². Large pieces of wood, leaves and stones, greater than approximately 1 cm, were removed by hand and sediments were air-dried at the room temperature. Dry sediments were homogenized, ground with a pestle and mortar and sieved using a 2 mm sieve. Samples of alluvial soils from river flood plains were taken from 14 sites by the standard procedure (Foster et al. 2000). Soils were sampled at 10×10 m sampling plots as 5-10 subsamples from 0 to 20 cm layer that were homogenized. Soil samples were quickly transported to the laboratory in polyethylene black bags and sieved through 2 mm mesh (with the exception of the portion for determination of the physical-chemical properties). All the sampled soils were fluvisols with four soil texture classes (loamy sand, sandy loam, loam, and clayey loam). Samples for chemical analyses were homogenized, air-dried at laboratory temperature and then sieved (2 mm), and stored at laboratory temperature.

Organic carbon content (Corg) and soil pHKCl were assessed by the standard methods described e.g. in Forster (1995). Particles below 10 μm were described as clay. For the analysis of organic pollutants, 20 g of dry sediment or soil were Soxhlet extracted for 8 h in dark with dichloromethane (DCM). The laboratory blank and the reference material were analyzed with each set of sediment or soil samples. Surrogate recovery standards (D8-naphthalene, D10-phenanthrene, D12-phenylene for PAHs analysis, PCB 30 and PCB 185 for PCBs analysis, 10 ng per sample) were spiked on each sample prior to extraction. Terphenyl and PCB 121 were used as internal standards for PAHs and PCBs analyses, respectively. Volume was reduced after extraction on rotary evaporator and under a gentle nitrogen stream at ambient temperature. Activated copper was used for sulphur removal. The fractionation was achieved on silica gel column; sulphuric-acid-modified silica gel column was used for PCB/OCP samples. Samples were analyzed using GC-ECD (HP 5890) supplied with a Quadrex fused silica column 5% Ph for PCBs (PCB 28, PCB 52, PCB 101, PCB 118, PCB 153, PCB 138, PCB 180, and OCPs (α-HCH, β-HCH, γ-HCH, δ-HCH, p,p’-DDE, p,p’-DDD, p,p’-DDT) and HCB. 16 US EPA polycyclic aromatic hydrocarbons were determined in all samples using GC-MS instrument (HP 6890 - HP 5973) supplied with a J&W Scientific fused silica column DB-5MS. Samples were quantified using Pesticide Mix 13 (Dr. Ehrenstorfer) and PAH Mix 27 (Promochem) standard mixtures.

Quality Assurance / Quality Control. Recoveries were determined for all samples by spiking with the surrogate standards prior to extraction. Recoveries were higher than 71 % and 69 % for all samples for PCBs and PAHs, respectively. Recovery factors were not applied to any of the data. Recovery of native analytes measured for the reference material varied from 88 to 103 % for PCBs, from 75 to 98 % for OCPs, from 72 to 102 % for PAHs. Laboratory blanks were very low.
KEY POLLUTANTS: SURVEY OF ENVIRONMENTAL CONCENTRATION LEVELS AND THE FLOODING INFLUENCE

This stage is focused on the description of potential or actual contact or co-occurrence of stressors with receptors. It is based on exposure measurements, as well as on the characteristics of ecosystem and receptor which are used to analyze stressor sources, their distribution in the environment, and the extent and pattern of contact or co-occurrence. This stage also describes the impact and uncertainty on exposure estimates and reaches a conclusion about the likelihood that the exposure will occur.

In the first place, it is very important to know what kind of contaminant sources is present in the study area, where these sources are situated, and which stressors are important for our endpoints (Problem formulation); for complete list of measured parameters see Table 2, only selected stressors were adopted for further analyses. The exposure assessment in this study arises from the environmental monitoring of the pollutants concentrations in our region. Layers displaying the measured concentration of selected stressors are shown here. These values represent the real contamination to which the receptors are exposed.

<table>
<thead>
<tr>
<th>Physico-chemical parameters</th>
<th>Chemical contaminants</th>
<th>Biological parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HMs</td>
<td>PAHs</td>
</tr>
<tr>
<td>pH (H2O)</td>
<td>Cd</td>
<td>Naphthalene</td>
</tr>
<tr>
<td>pH (KCl)</td>
<td>Cr</td>
<td>Acenaphthylene</td>
</tr>
<tr>
<td>Ntot</td>
<td>Hg</td>
<td>Acenaphthene</td>
</tr>
<tr>
<td>CaCO3</td>
<td>Ni</td>
<td>Fluorene</td>
</tr>
<tr>
<td>Corg</td>
<td>Pb</td>
<td>Phenanthrene</td>
</tr>
<tr>
<td>Q4/6</td>
<td>Zn</td>
<td>Anthracene</td>
</tr>
<tr>
<td>FA (fulvo acids)</td>
<td></td>
<td>Fluoranthene</td>
</tr>
<tr>
<td>HA (humic acids)</td>
<td></td>
<td>Pyrene</td>
</tr>
<tr>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ca</td>
<td></td>
<td>Benzo(a)anthracene</td>
</tr>
<tr>
<td>Mg</td>
<td></td>
<td>Chrysene</td>
</tr>
<tr>
<td>K</td>
<td></td>
<td>Benzo(b)fluoranthene</td>
</tr>
<tr>
<td>CEC (cation exchange capacity)</td>
<td></td>
<td>Benzo(k)fluoranthene</td>
</tr>
<tr>
<td>C/N ratio</td>
<td></td>
<td>Indeno(123cd)pyrene</td>
</tr>
<tr>
<td>Clay</td>
<td></td>
<td>Dibenzo(ah)anthracene</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Benzo(ghi)pyrenyl</td>
</tr>
</tbody>
</table>

Table 2) measured parameters in “Zlín” case study

The picture below displays the contamination to assess the presence of stressors before and after the floods in 1997. You can clearly see the changes in spatial distribution of the contaminants. A map with this data displayed will be one of the most important parts of our exposure assessment. The GIS technology enhances considerably our ability to view and analyze the environmental concentration of pollutants in relation to other
supplementary information. The number of other layers in this map project can be very useful for exposure assessment, too.

Because of soil assessment in this case, the layers describing pH, geology, pedology or land use can be very helpful to a good exposure specification. You can display the layers “pH”, “Geology” or “Land use” for these purposes. You can reach this information for our sampling sites, which will help you estimate exposure, as the soil properties play an important role in bioavailability and future conditions of the contaminants.
The analysis confirms the division of regions into 5 subregions with homogeneous pattern of contamination and basic characteristics (Table 2). The similarity within these regions was not driven solely by the actual concentrations but mainly by the trend of their changes (behavior) in time. Fig. 1 displays multivariate clustering of sites on the basis of concentration levels of all examined pollutants (PAHs, PCBs, OCPs, HCB). The clustering is intentionally conducted over the whole monitored time period and therefore comprises the time-related variability, including the influence of flooding. This analysis documents that contamination data relatively efficiently separated soil samples according to regions, while sediments data were substantially more
variable. As for soil samples, there is a less clear position of Region II which resembles to Region I (due to PCBs pattern) and to Region III (due to PAHs pattern), but other than that the soils are clearly separated into the regions. The cluster analysis for sediments allowed us to recognize clearly only specific position of Region I and III, while data of samples from other sites could not be distinguished. It is caused namely by major fluctuation of values after floods. To sum up, the data can be reasonably aggregated in the regions that reflect stratification of the area of interest according to environmentally important factors (Region I – V). The regions can be recognized on the basis of soil contamination, reflecting sources and the sample character, even if we ignore time-related fluctuation. The greater variability of sediment data document a less stable environmentally-related memory of sediments compared to soils. They show much more dynamic changes responding strongly to the actual situation, and they are affected by much subtle changes than soils, such as increase or decrease of flow.

<table>
<thead>
<tr>
<th>Contaminants</th>
<th>Region I.</th>
<th>Region II.</th>
<th>Region III.</th>
<th>Region IV.</th>
<th>Region V.</th>
<th>p value 1</th>
<th>Components of variability</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEDIMENTS</td>
<td>Median values (10 %; 90 % quantiles); calculated over the whole monitored period</td>
<td>Between region differences</td>
<td>Time-related differences</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Σ PAHs (mg.kg⁻¹)</td>
<td>10.5 *</td>
<td>23.7 *</td>
<td>11.2 *</td>
<td>13.4 *</td>
<td>7.6 *</td>
<td>0.023</td>
<td>28.6 %</td>
</tr>
<tr>
<td>(6.5; 17.6)</td>
<td>(10.8; 41.4)</td>
<td>(11.3; 21.7)</td>
<td>(4.3; 19)</td>
<td>(5.6; 36.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Σ PCBs (μg.kg⁻¹)</td>
<td>47.5 *</td>
<td>23.1 *</td>
<td>31.8 *</td>
<td>20.2 *</td>
<td>37.1 *</td>
<td>0.821</td>
<td>2.5 %</td>
</tr>
<tr>
<td>(4.3; 88)</td>
<td>(5.5; 92.3)</td>
<td>(6.9; 85.9)</td>
<td>(8.3; 72.6)</td>
<td>(7.3; 75)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Σ HCH (μg.kg⁻¹)</td>
<td>0.42 *</td>
<td>0.37 *</td>
<td>0.75 *</td>
<td>0.46 *</td>
<td>0.59 *</td>
<td>0.765</td>
<td>7.4 %</td>
</tr>
<tr>
<td>(0.14; 2.1)</td>
<td>(0.11; 2.45)</td>
<td>(0.25; 2.8)</td>
<td>(0.08; 1.6)</td>
<td>(0.08; 3.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Σ DDT (μg.kg⁻¹)</td>
<td>47.5 *</td>
<td>3.8 *</td>
<td>4.6 *</td>
<td>10.1 *</td>
<td>11.6 *</td>
<td>0.148</td>
<td>9.8 %</td>
</tr>
<tr>
<td>(1.6; 11.8)</td>
<td>(1.5; 22.9)</td>
<td>(1.8; 15.3)</td>
<td>(3.7; 49)</td>
<td>(5; 47.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HCB (μg.kg⁻¹)</td>
<td>1.6 *</td>
<td>1.3 *</td>
<td>1.9 *</td>
<td>5.2 *</td>
<td>3 *</td>
<td>0.032</td>
<td>19.8 %</td>
</tr>
<tr>
<td>(0.88; 2.5)</td>
<td>(0.47; 5.7)</td>
<td>(0.7; 3.5)</td>
<td>(2.3; 43.1)</td>
<td>(0.9; 7.5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOILS</td>
<td>0.4 *</td>
<td>14.9 *</td>
<td>11.9 *</td>
<td>8.1 *</td>
<td>12.2 *</td>
<td>&lt;0.001</td>
<td>52.4 %</td>
</tr>
<tr>
<td>Σ PAHs (mg.kg⁻¹)</td>
<td>(0.27; 3)</td>
<td>(2.25; 24.8)</td>
<td>(5.3; 15.6)</td>
<td>(1.4; 10.3)</td>
<td>(8.4; 16.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Σ PCBs (μg.kg⁻¹)</td>
<td>3.3 *</td>
<td>7.4 *</td>
<td>30 ab</td>
<td>37.7 b</td>
<td>40.8 b</td>
<td>0.002</td>
<td>33.6 %</td>
</tr>
<tr>
<td>(1.7; 17.8)</td>
<td>(2.7; 23.6)</td>
<td>(8.9; 84.8)</td>
<td>(14.6; 93.2)</td>
<td>(17.3; 91.2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Σ HCH (μg.kg⁻¹)</td>
<td>0.38 *</td>
<td>0.52 *</td>
<td>2.6 ab</td>
<td>0.83 ab</td>
<td>1.7 ab</td>
<td>0.023</td>
<td>38.7 %</td>
</tr>
<tr>
<td>(0.17; 1.2)</td>
<td>(0.25; 1.3)</td>
<td>(0.75; 8.5)</td>
<td>(0.38; 2)</td>
<td>(0.88; 3.4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Σ DDT (μg.kg⁻¹)</td>
<td>0.79 *</td>
<td>2.9 b</td>
<td>10.3 b</td>
<td>4.6 b</td>
<td>8.7 b</td>
<td>0.019</td>
<td>8.6 %</td>
</tr>
<tr>
<td>(0.51; 3)</td>
<td>(1.4; 36.5)</td>
<td>(3.4; 55.1)</td>
<td>(2.5; 42.9)</td>
<td>(4.1; 22.6)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HCB (μg.kg⁻¹)</td>
<td>1.2 a</td>
<td>2 a</td>
<td>3.9 b</td>
<td>3.6 ab</td>
<td>3.2 ab</td>
<td>0.016</td>
<td>13.2 %</td>
</tr>
<tr>
<td>(0.39; 2.2)</td>
<td>(1.2; 4)</td>
<td>(1.2; 35.3)</td>
<td>(2.6; 9)</td>
<td>(1.9; 8.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 Overall p value of Kruskal-Wallis test comparing regions
2 Components of overall variability that belong to the differences between regions and to Time-related differences (trends and fluctuations). Calculated as ratios of relevant sum of squares (ANOVA model; based of log-transformed concentration data).
3 Marks of statistical significance of multiple comparison tests between regions. Values within one row marked by the same letter are not mutually significantly different (p > 0.05; multiple median test)

Table 3) Statistical description of sub-regions
The data from our study document that redistribution from the contaminated sediments to floodplain soils can also play an important role, namely for PAHs. Our study documents that flood events have a major impact on transport of particle-bound pollutants. Dominant role of floods on the development of floodplains in the regularly flooded areas and the significance of floodplains as a valuable stratigraphic record of the past floods has been previously documented (Benedetti 2003). There are two possible sources of contamination to alluvial soils during floods: 1. remobilization and redistribution of contaminants from sediments, 2. contamination through flood water from material and pollutants that have been mobilized from flooded contaminated terrestrial areas including industry, landfills, old loads or sewage treatment plants. In the second case, sediment contamination could also increase. The deposition of nutrient and contaminant rich suspended particulate matter originating from sediments has been identified as an important factor in high contamination of recent floodplain areas along river Elbe (Forstner et al. 2004). Similarly to our study, the risks in the Forstner’s study have been connected with the deposition of contaminated solids on soils, with floodplain soils representing a record of the pollution history in the catchment area (Forstner et al. 2004).

All the studied soils were flooded in 1997 and stayed under water for more than three weeks. The floods strongly affected the soils characteristics. Range of values reflects the variability in abiotic conditions in soil environment throughout the monitoring period. Major source of within site fluctuations was due to flooding in 1997. pH conditions appear to be relatively stable, while total organic carbon and clay content changed significantly due to water and sediment disposal – that is why the soil texture categories were assessed in the stable conditions (1-2 years after water disposal).
A. Sediment data

B. Soil data

Dates of sampling campaigns (month-year)

Figure 6) Time related changes of contaminants in sub-regions

Strong fluctuation and disposal of new material during floods also influenced the correlation of organic carbon level with contamination level. The indication potential of correlation patterns was therefore strongly diminished due to the influence of floods. Analysing sediment samples, we found positive relation only between PAHs and Corg content, more pronounced in the case of high MW PAHs. However, this correlation
was strongly influenced by similar time trends in Corg content and PAHs and similarly to PAHs (Fig. 3), organic carbon was temporarily removed due to floods. Therefore, the Corg content revealed the opposite time trend than PCBs and the negative correlation between PCBs and Corg in sediments shown in Table 4 is an artifact related to the flooding. An important role of organic carbon in total sorption capacity of sediments for organic compounds has been documented in previous studies (Chiou et al. 1998). In some field studies, strong correlations were found namely between the higher MW PAHs and Corg content (Evans et al. 1989), but there is a number of field studies where the relationship between the organic carbon content and the PAH concentration is ambiguous. The absence of correlation is sometimes considered as an indicator of anthropogenic pollution; it has been documented that sites with greater level of contamination distort the correlation relationship and can cover the influence of Corg (Evans et al. 1989). Our data confirm that the relationship to the organic carbon is stronger for higher MW PAHs, but also stress the issue that it can be significantly distorted by flooding events.

KEY POLLUTANTS: LOCAL HOTSPOTS AND ZLÍN URBAN AGGLOMERATION

These results of the analysis confirmed the importance of flooding for contamination of soils. Moreover, some hotspots were identified in the region: some of them are industrial sources, other are diffusion sources.

Many cases of risk substances releases were found, among the most hazardous sources belong Barum Continental (rubber industry – release of PAHs, PCBs, Cd, Cu, Pb, Zn) and Hamag (foundry – release of PCBs, HMs, PAHs) in Zlín – Otrokovice agglomeration. Both companies were flooded and washed out during the floods in July 1997.
There is also influence of different types of landscape, especially Zlín agglomeration on levels and combinations of environmental pollutants.
Figure 8) The example maps of the regions contamination with chosen pollutants (sum of PAHs, sum of PCBs, sum of DDTs)
CONCLUSIONS

The sediments showed the potential to serve as a secondary source of contamination for the aquatic ecosystem, but also for floodplain soils and other flooded areas. The floods caused significant decrease of all contaminants in sediments; on the other hand, concentration of PAHs significantly increased in alluvial soils. PCBs were washed out from the region, while PAHs have moved within the region. The remobilization of contaminated river sediments, subsequently deposited on the inundated areas, probably resulted in an increase in soil contamination with PAHs; therefore, floodplain soils represent a significant sink for these compounds. The sediments show the ability to reload the fresh sedimented material with pollutants from continuous contaminant sources, predominantly combustion processes, after having been washed out during floods. The reloading of river sediments in time, namely with PAHs, increases their risk as a potential source in the next flood event both to downstream sediments and affected alluvial soils. Some of the affected alluvial soils are used as grassland to produce feed for livestock and the pollutants from alluvial soils could be potentially taken up by these animals and enter the food chain with possible risk for humans. Therefore, it is important to pay attention to the floodplain soil contamination and include them into the risk assessment focused on flood effects. Soils have a much stable long-term environmental memory related to contamination levels, pattern and distribution, so they can provide more relevant information on the overall contamination of the area. On the other hand, sediments are much more dynamic, reflecting subtle changes; the immediate situation within the watercourse can therefore significantly affect the results of the analysis and lead to an underestimation or overestimation of the general state of contamination.

Significant hot spot sources were identified - they release large volume of the monitored contaminants. Their influence on the soil and sediment contamination was proved.
III. EFFECT ASSESSMENT

RELEVANT RECEPTORS

There are two main approaches for biological receptors in EcoRA. The first one is the application of ecotoxicological tests for the measurement of influence of doses of environmental pollutants: this method can be applied on any level of EcoRA.

The other approach is the ecosystem monitoring using biological communities, which can also be used on any level of EcoRA, but to answer different questions.
The following text describes the selection of organisms for ecotoxicological tests; the biomonitoring network is described in a separate chapter.

The main aim of this study was to evaluate ecological condition of the alluvial soils. Therefore, we chose such receptors and observed endpoints which are important for soil processes and soil quality. The effects are assessed on the Poaceae, soil microorganisms, the Lumricidae, the Enchytraeidae and the Isostomidae.

From the family Poaceae, Avena sativa was selected as a assessed species because of its frequent usage in ecotoxicological tests. Therefore, a sufficient amount of toxicological data from databases and literature for different contaminants is available. As an arable crop, it is not exactly a representative species for inundated soils; this will be taken into account by next steps.

Selected endpoints are the growth and the mortality. PNEC values are from EC50 values derived.

For soil microorganisms, the data related to the functioning of microbe-mediated processes and/or enzymatic activity are used for the derivation of terrestrial PNEC (PNEC process)

The survival and the reproduction of Eisenia fetida (Lumbricidae) are the next assessed endpoints. This earthworm is not a highly relevant species for our samples - this problem will be taken into account by PNECs derivation from NOEC values.

From the family Enchytraeidae, Enhytraeus crypticus was selected as a relevant species for our soils. The assessed endpoint is the reproduction and the PNEC values will be estimated from available NOEC values for selected pollutants.

The number of young individuals is the endpoint for in ecotoxicology often used collembolan Folsomia candida. PNECs are derived from NOEC values.

If NOECs for individual species of more taxonomic groups are available, derivation of PNEC species can be used. Or we can find out these values from some databases.
BIOMONITORING NETWORK: REFERENCE BASED APPROACH FOR EFFECT ASSESSMENT

The analysis of biomonitoring networks and comparison of unknown localities towards reference sites is a promising approach for EcoRA analysis of evidence of impacts in the past, benchmarking for the status of valuable ecosystems and long term and wide scale changes in the environment.

There are several possible application of this approach in this studied area: one of them is connected also to the implementation of the Water Framework Directive in the Czech Republic and works with the communities of benthic macroinvertebrates and their links to environmental conditions. The European Water Framework Directive (Directive 2000/60/EC) requires a good environmental condition of the water bodies in the member states of the EU and this demand, of course, requires the methodology of evaluation of ecological condition of waters, i.e. the EcoRA methodology.

Monitoring of water organisms communities has become a standard approach in surface water monitoring, as well as a part of complex systems to assess surface water quality. In European countries, the most commonly used organisms are water macroinvertebrates; developed in the United Kingdom (Wright et al., 2000; Clarke et al., 2003), RIVPACS has been one of the first complex systems based on macroinvertebrates. Similar systems have been used in many countries worldwide (Barbour et al., 1999) and development of similar systems is also connected to the European Water Framework Directive (Logan and Furse, 2002).

DATA SOURCES

The correct data are crucial for any analyses and development of standardized methodology for the evaluation of the ecological status of surface waters. According to WFD EU the following set of data is necessary:

- data on communities of benthic macroinvertebrates, fishes, phytobenthos, phytoplankton and macrophytes;
- Supporting data of chemical monitoring and hydromorphology.

The data should contain structure and composition of biological communities for all biological compounds, properties of taxa, i.e. species traits and influential abiotic factors (natural parameters, altitude etc.) and stressors. All these data should be available for both reference and contaminated sites.

BIOLGICAL COMMUNITIES AS ASSESSMENT ENDPOINT IN THE ECORA

The biological communities and their biodiversity have several substantial advantages in the EcoRA process:

- Long-term „memory”
- Interpretation at ecosystem level
- Relation to functionality and stability of ecosystems
- End-point with evolutionary background

However, there are also some serious drawbacks:

- Time-consuming measurement
- Problem with representative sampling
- Dependent on experts (taxonomists, ...)
- Long-term representative data are required for benchmarking
- Diversity = variability: numerous masking effects
In addition, EcoRA requires a correct statistical methodology for analysis of biodiversity. The biodiversity and properties of the biological communities is the key concept of ecology and its understanding and analysis is crucial for any ecological research. The first possible problem in this process is the term “biodiversity” itself because of a different comprehension of its meaning. The diversity definition range from the simplest definition of biodiversity as the number of species (in fact, species richness (Fiedler and Jain, 1992) to complex definitions of DeLong (1996): “Biodiversity is an attribute of an area and specifically refers to the variety within and among living organisms, assemblages of living organisms, biotic communities, and biotic processes, whether naturally occurring or modified by humans. Biodiversity can be measured in terms of genetic diversity and the identity and number of different types of species, assemblages of species, biotic communities and biotic processes, and the amount (e.g., abundance, biomass, cover, rate) and structure of each. It can be observed and measured at any spatial scale ranging from microsites and habitat patches to the entire biosphere.”

The latter definition covers the whole complexity diversity and brings up the question: How to analyze such a complex system? The set of methods should be applied to cover all aspects of biodiversity and the main aim of this paper is to present the main approaches for diversity analysis and the advantage of their combination.

STATISTICAL METHODOLOGY
The main idea of the analytical system is an approach based on a network of reference sites and combination of multimeteric approach and robust multivariate modelling of expected environmental and biological conditions (could be called “RIVPACS type” from the first system using this methodology, developed in the United Kingdom).

As for the analysis of surface water quality based on biological communities, there are several methodology approaches: single metrics [5], multimetric [3], [6] and a multivariate approach [1]. All these metrics were implemented on different levels of computation. On this place the short introduction to application of statistical methods on biological communities in EcoRA is required.

APPROACHES IN ANALYSIS OF BIODIVERSITY DATA
As stated above, the biodiversity data can be analyzed from many points of view (Fig 1).

Figure 1) Biodiversity and its analytical questions.

There are four main approaches to answer these questions; these approaches can be applied on any type of diversity (taxonomical, genetic, functional etc. level):

- Diversity indices
  - The whole biological community is aggregated into one number which represents the number of species and/or dominance of species in the community
  - The confidence intervals can be obtained for this type of computation and tests for indices comparison are available; the biodiversity indices can be considered as an analogy to descriptive statistics
The advantage of diversity indices is that they provide measure of diversity in one number; their disadvantages are reduction of species individuality and sometimes unclear interpretation of indices (the same value of index can occur in very different communities)

- **Biotic indices**
  - These measures (saprobi index is probably the best known) combine knowledge of species traits obtained from the laboratory experiments, expert knowledge or analysis of valence curves with communities structure and provide measure of functional diversity of the community
  - The main advantage is the incorporation of additional information on species, nevertheless it is also its biggest disadvantage because these data are sometimes lacking or suffer from uncertainties

- **Species abundance models**
  - The quantitative structure of biological community (species abundance profile) results from the underlying ecological processes and the model profile with possible ecological interpretation can be fitted to profile of observed communities
  - This approach can be considered as analogy to distribution fitting in statistics
  - The advantage is that the whole abundance profile of community is included in the analysis (which is lacking in diversity indices) and that the species are represented only by the ranked abundance profile (i.e. the hypothesis of ecological process influencing quantitative structure of community is tested instead of abundance profile of individual species); the disadvantages of this approach are unfinished methodology of fitting of expected and observed abundance profiles and lack of information on testing of ecological hypothesis connected to model profiles

- **Multivariate analysis of biological communities**
  - The individual species and their quantification in communities can be subject to many types of cluster, ordination, regression and classification multivariate techniques.
  - In contrast to previous methods, which can be computed for single community, these methods search for multivariate patterns in huge datasets of communities and can answer questions about the relation of species to environment, their spatial relationships and species interactions. Moreover, results from the first two approaches can be used as an input for multivariate data analysis.
  - The advantages of multivariate methods are comprehensive results aggregating all aspects of the data and possibility to reveal hidden interactions and links among variables; disadvantages are that these methods are data demanding and rather complicated and require expertise both in their statistical methodology and biological communities under study, otherwise can lead to incorrect conclusions and interpretations.

All types of analyses can be applied for various types of biodiversity i.e. taxonomy, genetic or functional biodiversity.

**Biodiversity sampling**

Similarly like in other environmental sampling, we lack the resources for complete sampling of communities (with some exceptions like parasite infracommunities etc.) and the observed species are a result of ecological
processes and sampling bias. The most important parameters are sample size, spatial and temporal variability. The optimal sample size can be estimated from screening studies (Fig. 2), the spatial and temporal variability should be eliminated by sampling design when taking possible ecological processes into consideration.

![Image](image-url)

Figure 2) Influence of sample size on number of species (Jarkovský & Brabec, 2004)

There is a set of methods for estimation of species richness in sample of given size; it allows us to compare species richness even in unequally sampled samples. The best known is rarefaction method (Krebs, 1989). Given a column of abundance data for a number of taxa, it estimates how many taxa you would expect to find in a sample with a smaller total number of individuals. With this method, you can compare the number of taxa in samples of different size. Using rarefaction analysis on the largest sample, you can find out the number of expected taxa for any smaller sample size.

**Diversity indices**

The diversity indices measure various aspects of biological communities in one aggregated value and are based on empirical approach or mathematical theory. The indices can be separated in two groups:

- Indices measuring number of species
- Indices combining both number of species and their relative abundance; dominance indices are also part of this group

The simplest index from the first group is species richness (number of species), the others are Margalef (Cliphord & Stephenson, 1975) and Menhinick (Whittaker, 1977) index which combine number of species and sample size. The most useful from this group is the simple species richness; nevertheless, it should not be used out of context of sample size and other diversity measures.
The latter group of indices is quite heterogeneous: there are indices based on information theory like Shannon (Pielou, 1975) and Brillouin index (Pielou, 1969, 1975) which combine both number of species and equality of species abundance – the high value of index means high number of species and equal abundance of species. The indices have maximum possible value for given number of species and thus the relative index called evenness can be computed. The other group of indices is dominance indices; some of them – like Simpson index (Simpson, 1949) – are highly correlated to previous group and in fact did not bring new quality. Other are based on different principles – like Berger-Parker index (Berger & Parker, 1970, May, 1975), which compares the most abundant species to sample size. Another group, often correlated with Shannon index, comprises indices derived from the species abundance models like alpha of logarithmic series (Fisher et al., 1943) or gamma of log normal distribution model (Preston, 1948, 1962). Another example of these indices is Q statistics (Kempton & Tailor, 1976,1978) and its stochastic form (Dušek et. al. 1998) which can be used as the replacement of Shannon index with better statistical properties; it is derived from the abundance curve of cumulative number of species.

There are dozens to hundreds of various diversity indices, but their usefulness is in doubt because of their high correlation with each other (Fig 3); an extensive list of indices is given in Magurran (1983) or Harper (1999). We can conclude that the diversity indices should be used in a combination which reflects different aspects of biological communities, i.e. species richness, dominance and evenness of abundance: the selection of one index from each group is sufficient for the further analysis.

Figure 3) Correlation of diversity indices (Jarkovský & Kubošová, 2007, unpublished)
VARIABILITY AND TESTING OF DIVERSITY INDICES

As stated above, the diversity indices can be considered as descriptive statistics of biodiversity, nevertheless their confidence interval and statistical testing is far more complicated than in common statistics, like mean or standard deviation. Three approaches can be applied: (1) for some indices, approximation to t statistic can be computed (Hutcheson, 1970) and adopted for the comparison and confidence interval estimation, (2) empirical distribution is provided for some indices like stochastic version of Q statistic (Dušek et. al. 1998) and standard non-parametric tests can be applied, and (3) jack-knife or bootstrap estimators can be applied for confidence interval computation and testing of any diversity index; this approach is implemented in the PAST software, for example.

SPECIES ABUNDANCE MODELS

The species abundance models represent an approach trying to categorize biological communities according to defined abundance profile with possible biological interpretation, i.e. the possible biological interpretation is important advantage of this method. The models are derived from the common mathematical function, and biological interpretation is assigned to this function; or the models are defined by ecological theory and statistically described by compute simulation. The latter models have a stochastic nature and although they are promising especially for small communities, many questions still arise about their fitting to observed data. In this sense, the models can be divided into statistical and biological models with special subgroup niche oriented models (Tokeshi, 1993). Although the models have different origin, they can provide similar results. This is documented on Fig 4, where both statistical and biological models are displayed. Similarly as for diversity indices, several general types of models can be selected here as well: from model with high dominance to equally distributed species (and several intermediate types, of course).

Figure 4) Comparison of species abundance models (based on Tokeshi 2003)

The methodology of comparison between the observed and predicted community pattern has not yet been finished. The simplest (but rather subjective) method is the optical comparison of curves; the statistical
methods range from goodness of fit test (suitable only for large communities), Kolgomorov-Smirnov test or Hellinger distance (Tokeshi, 1993) or simulation techniques and Monte Carlo approach (Bersier, Sugihara 1997).

**MULTIMETRIC COMPARISON WITH REFERENCE CONDITION**

The core statistical analysis for the indices combination is based on the standardized methodology of multimetric indices development (Hering, 2006), which consists from the following steps. Firstly, the list of candidate metrics is prepared; the metrics adopted for this project are biodiversity measures (Magurran, 1983), saprobity indices (ČSN 75 7221), species traits scores (WAES: weighted abundance ecological score; WSES: weighted species ecological score) based on species traits from the AQEM project (Hering et al., 2001, Sandin et al., 2000, 2001), and various hydrobiological indices adopted from the ASTERICS software (ASTERICS 2.3). Both specialized computational library developed for the project and ASTERICS software were adopted for the computation.

The following steps are aimed on the selection of indices with the clear relationship to condition of the water bodies. The first step is the selection of the appropriate sets of indices (metrics) which reflects gradients of different stressors (i.e. can distinguish between reference and non-reference condition) and here are selected by the means of Mann Whitney U test. This analysis is followed by the correlation analysis (Spearman correlation) of metrics and environmental variables which define river types, i.e. the metrics should reflect stressors and should be predictable by environmental used for the definition of river types. Because many indices are redundant, it is necessary to find the sets of indices which are independent on each other, i.e. the metrics are organized into modules; shortened list of indices is selected from the each module for the final computation. This is done by means of the cluster analysis. The selected list of metrics is then validated against aggregated time series of the measured environmental stressors (Table 1).

<table>
<thead>
<tr>
<th>Reference sites (N)</th>
<th>Non-reference sites (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biochemical consumption of oxygen</td>
<td>157</td>
</tr>
<tr>
<td>Consumption of oxygen by dichromate</td>
<td>157</td>
</tr>
<tr>
<td>Total organic carbon</td>
<td>152</td>
</tr>
<tr>
<td>Conductivity</td>
<td>157</td>
</tr>
<tr>
<td>Ammonia nitrogen</td>
<td>157</td>
</tr>
<tr>
<td>Nitrite nitrogen</td>
<td>157</td>
</tr>
<tr>
<td>Nitrate nitrogen</td>
<td>157</td>
</tr>
<tr>
<td>Total phosphorus</td>
<td>157</td>
</tr>
<tr>
<td>Total nitrogen</td>
<td>119</td>
</tr>
<tr>
<td>Mercury</td>
<td>152</td>
</tr>
<tr>
<td>Cadmium</td>
<td>152</td>
</tr>
<tr>
<td>Lead</td>
<td>152</td>
</tr>
</tbody>
</table>

Table 1) List of measured stressors on sufficient number of sites

The results of these steps are the so called modules of the final multimetric; each of them is reflecting different stressor and they are put together using “one out – all out principle”. These “indices” modules are combined with the main module reflecting general environmental degradation and based on the multivariate comparison of biological community composition with the model of reference quality (Fig. 5).
Above the direct comparison with expected natural condition, there is the multimetric combination of partial measures, forming the final measure of the ecological condition.

![Diagram of Environmental gradient and Model of reference state](image)

**Figure 5)** Basic principles behind the indices evaluation of ecological status (stressor gradient – index) and direct comparison of expected (reference based) and observed status.

The results of the computation are expressed as EQR (Environmental Quality Ratio), i.e. the ratio between expected values (based on reference data, i.e. data with good ecological status) and observed values (i.e. values of evaluated sites and water bodies); EQR=observed/expected. The final measure is expressed in range 0-1 and categorized into five classes of the ecological status.

**MULTIVARIATE PROBABILISTIC COMPARISON WITH REFERENCE STATUS**

These “indices” modules stated above are combined with the module reflecting general environmental degradation and based on the multivariate comparison of biological community composition with the model of reference quality. Above the direct comparison with expected natural status there is the multimetric combination of partial measures which forms the final measure of ecological status for the EcoRA process.

This very important module (RIVPACS type) has the very simple theoretical background, nevertheless its implementation presents a difficult task. The process could be divided into three steps, each of them with its own methodological problems: (1) preparation of the reference model, (2) classification of unknown cases (sites) into reference categories, and (3) comparison of an unknown site and the reference status, i.e. assessment of unknown site quality. The principal basis of the whole analysis is the quality of the reference dataset, which should homogeneously cover all environmental conditions in the analyzed area and contain minimal influence of human activities. Some guidelines for selection of reference localities are given in Hughes, 1994. Unfortunately, in the conditions of Central Europe, it is almost impossible to find real “natural” sites; thus, localities in the reference dataset originate from long-term knowledge of the considered sites and consensus of hydrobiology experts on what is the site’s “nearest-to-natural-conditions” status.
According to an ecological theory stating that the composition of biological communities is highly influenced by the environment, a link may be identified between standard and reference sites to see, whether under the same environmental parameters, these sites have got the same or different community composition. In other words, is there a shift in community composition in the case of a standard locality in comparison to the "natural" (reference) composition of biological community under the same environmental conditions (in the sense of hydro-morphological or hydro-geological properties of the respective sampling site)? The usage of reference datasets in evaluation of expected natural conditions consists of several steps (Fig. 6).

First of all, we have to define a homogeneous group within the reference database based on comparison of biological communities’ composition, i.e. to define a reference model consisting of several homogeneous categories according to their community composition. Hierarchical agglomerative clustering on distance matrix of biological communities followed by algorithm for definition of optimal number of clusters and expert opinion will be used for definition of reference groups. The definition of homogeneous groups of reference clusters of reference data, i.e. sites with similar composition of biological communities consists of several steps including both statistical methodology and expert judgment. The defined clusters were than validated by the silhouette algorithm (Rousseeuw, 1987) and expert judgment.

All analyses were computed for community of species with occurrence on more than 10 sampling sites, for the community of indicative species and EPT taxa; this brings another problem and it is the selection of the set of appropriate indicative species. Combination of two approaches i) Random Forests method (Breiman, 2001) and logistic regression (ter Braak, 1996) was adopted for the selection of species for the further analysis.

These groups should be defined by hydro-morphological parameters of the respective sites, i.e. by parameters which should not be influenced by human activities; these parameters are used for classification of unknown sites (standard monitoring) into reference groups according to theory that sites with similar environmental conditions should have similar biological communities. For the comparison of observed and predicted conditions the natural environmental status should be based on reference groups; the natural conditions should be defined for biological communities, environmental metrics and chemical pollutants. The sites of monitoring network with unknown ecological status should be classified into reference categories. There are a
The simplest and the most objective measure of object association in multivariate space is their distance; thus, we decided to build our method on an analysis of a distance matrix among localities. Now, selection of proper distance metric is the first task in designing the method. We have adopted Gower metric (Gower, 1971); however, any multivariate distance metric suitable for given data could be used. Concerning biomonitoring data, there are some advantages in Gower metric: i) Continuous, binary or categorical parameters may be incorporated in computation: binary data is computed by coefficient – agreement and disagreement of values forming distance 0 or 1 respectively; categorical data is computed in the same way. Distance of objects according to continuous data is weighted to i) a parameter range in the data file or ii) an externally provided parameter range, i.e. difference in parameter values of objects is divided by parameter range to obtain partial metric ranging from 0 to 1. ii) As noted above, parameters are weighted to their range, i.e. the influence of parameter absolute value is removed. iii) The final distance metric ranges from 0 to 1 and could be easily interpreted. iv) Parameters in computation could be weighted according to expert knowledge or results of preliminary analysis. The final metric takes the following form:

\[ D(x_i, x_j) = \frac{\sum w_j d_{ij}}{\sum w_j} \]

where D is distance between objects x1 and x2, d1j is partial distance of objects x1 and x2 associated with parameter j (there are 1..p parameters; partial metric associated with parameter ranges from 0 to 1) and wj is weight of parameter j with range 0-1.

Every homogeneous category of reference data could be characterized by its position in the multivariate space; and also, by its multivariate variability. Position of the reference category centroid (based on the median of continuous data and modus of binary/categorical data) exhibits representative of this group; multivariate radius of group provides the measure of its variability (in fact 95% percentile of radius is used in our computation to remove the influence of outliers). The distance of an unknown case to the centroid (D) is compared to the percentile of the reference category range (R). This ratio measures the extent to which an unknown case differs from objects incorporated in the reference category. Due to the fact that reference categories are not probably multivariate spheres we had to add a safety measure reflecting the real multivariate shape of the reference data. There are two parameters incorporated in the computation: the distance of an unknown case to the nearest neighbor in the reference group (N) and the measure of intragroup distances (I) within the reference group. The measure of intragroup distances is taken as median length of the MST branches (minimal spanning tree; Prim, 1957) of objects in the reference group. The following formula gives the measure of distance of an unknown case to the reference group x (Ux) in multiplies of the reference group x radius weighted for multivariate shape of this group.

\[ U_x = \frac{\text{abs}(D + N - I)}{R} \]

This computation could be also expressed as a probability of case U belongs to group x: \[ P(U_x) = \frac{1}{U_x} \times 100 \] where values over 100% (i.e. objects inside the reference group) are truncated to 100%. In the first step of the analysis (classification of an unknown case into reference groups according to natural heterogeneity), P(Ux) is computed for all
reference groups \( x=1..n \) and probability of unknown case belongs to a particular group is weighted as follows

\[
P_W(U_x) = \frac{P(U_x)}{P(U_1) + P(U_2) + \ldots + P(U_n)}.
\]

The output of the classification method is the probability of assigning a locality into the reference class based on natural heterogeneity, i.e., to which reference class the evaluated locality belongs; these probabilities are crucial for further evaluation of ecological status. As already mentioned, the comparison of predicted and observed ecological status is computed differentially for i) biological communities, ii) biotic indices and environmental parameters and iii) complex measure difference in multiple indices and/or environmental parameters between reference dataset and analyzed site. For the comparison of observed and predicted community composition, the RIVPACS approach for prediction of biological community composition, which incorporate three parameters (\( S_{gz} \): weight of reference group \( z \) according to its size; \( P(x)gz \): probability of classification of unknown site into reference group \( z \); \( pSp\text{c}(y)gz \): probability of occurrence of species \( y \) in reference group \( z \)) in computation of probability of occurrence of all species sampled in the monitoring (5):

\[
pSp\text{c}(y,x) = \frac{S_{g1} \times P(x)_{g1} \times pSp\text{c}(y)_{g1} + \ldots + S_{gn} \times P(x)_{gn} \times pSp\text{c}(y)_{gn}}{S_{g1} \times P(x)_{g1} + \ldots + S_{gn} \times P(x)_{gn}}
\]

The next step of the computation is the cut-off of species with low probability of occurrence (the threshold value is according to experiences of RIVPACS and similar systems set to 50% probability of occurrence). The expected number of species is computed as

\[
S_{\text{max}}(x) = \text{round}\left(\sum_y I_{[pSp\text{c}(y,x)<0.5]} \times pSp\text{c}(y,x)\right).
\]

The difference between observed and predicted status is computed. The difference is non-negative, therefore the predicted status is considered if it is greater than the observed status. Whereas the prediction of species abundances which are important for computation of biotic indices is connected with high level of noise and low precision (Guisan and Zimmermann, 2000), the prediction of expected values of biotic indices and also chemical parameters is based on another approach. The distribution of predicted values is based on frequency tables of values in reference data, where the probability of classification into reference group is used as the frequency weight of the given value. The predicted value is computed as median of predicted distribution and compared to observed value.

The final goal of the system is the measure of environmental disturbance of sites of monitoring network; it is defined as difference between observed ecological status of site and its expected status defined by probability of site classification into reference categories and the description of natural conditions based on reference dataset. The evaluation of status is computed for different partial measures of ecological status: Similarity measure (Jaccard coefficient) between predicted and observed composition of biological community and the Ecological Quality Indices (EQI) between measured value and median of predicted distribution of chemical pollutants or environmental metrics. The EQIs, according to RIVPACS methodology (Wright et al., 2000), are defined as the ratios of observed to expected values (\( O/E \)) of each parameter being used. The intention of using such ratios is that it provides a means of standardizing the biotic indices, so that a particular value of the EQI ratio implies the same ecological quality for that index, no matter what the type of site.

The last step of the analysis is combination of partial measures of ecological status into one final measure of ecological status which could be categorized into five categories according to WFD EU (from natural to most disturbed ecological status) and used for environmental reporting and management of ecological risk.
RESULTS

SELECTION OF CANDIDATE METRICS

The analysis was computed for the 317 reference sites and 365 non-reference sites and from the 105 candidate metrics most of them with exception of WSES - feeding type: xylophagous, WAES - microhabitat POM, WAES - zonation eukrenal, WAES - feeding type: xylophagous, WAES - feeding type: shredders, WAES - locomotion type: sprawling/waking, exhibit statistically significant difference between reference and non-reference sites.

The parameters with the highest importance for the river typology and unaffected by human activities were selected for the analysis (altitude, stream order) of correlation with 101 selected metrics with difference between reference and non-reference sites. 62 metrics with statistically significant correlation with altitude and/or stream order were selected for the further analysis. The analysis revealed three main groups of indices (modules): saprobity, diversity and habitat degradation; the final metrics were selected within these groups.

The metrics selected in the previous steps (22 metrics) were correlated with available stressors (Table 1). Margalef index and Alpha index were not correlated with any stressor; no metrics were correlated with heavy metals. Nevertheless the heavy metals are known as highly dangerous pollutants this can be explained: i) heavy metals are not common in toxic concentrations and ii) influence of heavy metals on benthic macroinvertebrates communities is low in comparison to oxygen content in water and saprobity.

PREDICTION OF MACROZOOBENTHOSES COMMUNITIES

The selection of indicative species is necessary step in the analysis and it removes the noise given by tolerant species with no indicative power. Analysis was computed using presence/absence data and only for taxa with presence on more than 10 sampling sites. Environmental parameters were selected using factor analysis of data and the altitude and stream order were selected as the most important and independent variables. The computation by random forests method provides importance parameter (with range 0-1) which gives us the information on indicative power of the species for given environmental parameters. Due to continuous nature of the species regression trees were adopted

As a supplementary approach for the definition of indicative species the valence curves were computed by the means of logistic regression. Tolerance parameter, the highest probability of species occurrence and optimum of environmental parameter were used for the definition of indicative power of the species towards the environmental gradient. The problems with this approach occur when the optimum is outside the range of analyzed environmental gradient and sigmoid or linear logit curve is used instead Gaussian logit curve.

From the total number of 1469 taxa we selected 56 taxa with indicative power for at least one environmental parameter. Clustering of reference sites according to Fig. 3 was confronted with expert opinion (Fig. 5) and the clustering based on Bray Curtiss distances was selected for the further analysis. In the next step the optimal number of clusters was selected by means of combination of expert opinion and silhouette metric. The resulted 20 clusters represent homogeneous categorization of reference data based on benthic macroinvertebrates communities and also correlated with environmental parameters not influenced by human activities (Fig. 6); the model is able to predict community composition on the basis of abiotic parameters and compute the difference between expected and observed communities applicable for the evaluation of the ecological status.

CONSTRUCTION OF THE FINAL MULTIMETRIC INDEX FOR THE EVALUATION OF THE ECOLOGICAL STATUS

The procedures described above lead to selection of 20 indices which formed the components of the final multimetric index. This index contains four modules:

- Saprobity: Saprobity index, RETI
• Diversity: Number of taxa, Number of individuals, Q index of diversity (Dušek et. al., 1998), Number of EPT (Ephemeroptera, Plecoptera, Trichoptera) taxa, Number of Plecoptera taxa

• Habitat degradation:
  o Zonality: zonation hypocrenal (WSES), zonation epirhithral (WSES), zonation epipotamal (WSES), zonation metapotamal (WAES), zonation hypopotamal (WAES), zonation litoral (WAES), zonation profundal (WAES)
  o Microhabitat preferences: microhabitat psammal (WSES), microhabitat pelal (WAES), microhabitat argyllal (WAES), microhabitat lithal (WAES),
  o Feeding preferences: feeding type: grazer and scrapers (WAES), feeding type: active filter (WAES)

• General degradation: based on separate predictive model

Due to fact that all metrics have different scales and/or correlation to gradient of sites degradation, it is necessary to recode them into the same scale and direction. This lead us to the set of transformations into EQR values which respect the value of parameter in reference network, range of parameters (both reference and non-reference sites) and direction of difference between reference and non-reference sites. The only difference from this concept is the general degradation of sites which is already produced as standardized EQR value.

CONCLUSION

The developed multimetric evaluation of the ecological status based on benthic macroinvertebrates is now under testing in routine usage and its implementation in the information system started. Multimetric indices like in our study provide a valuable tool for assessing various types of freshwater ecosystems, since they integrate different stressors and different components of the community; this approach allows us to evaluate the ecological status of surface waters and thus provide the tool required by the WFD EU for the decision processes and management of water bodies.

The problems of monitoring of surface waters, its analysis, reporting and interpretation reports have got the crucial role in the water management and planning according to WFD EU; for example in remedial processes, their control and also investment in this field.

Our current tasks according to WFD EU are the development of informatics solution of the methodology, the analysis of sufficient level of taxonomical determination of biological communities for routine biomonitoring (cost vs. gained information on ecological status), uncertainty analysis and to build up the reference model for biological communities utilizable in the routine biomonitoring.

ACKNOWLEDGMENT

The project is supported by the Ministry of Environment of the Czech Republic (project ARROW) and by project INCHEMBIOL (MSM0021622412).

REFERENCES


ČSN 75 7221: Czech State Norm for the Quality of Surface Waters.


ECOTOXICOLOGICAL TESTS: EVIDENCE-BASED APPROACH

INTRODUCTION

We can find dose-response experiment in the wide range of biological branches. Especially, majority of the toxicological experiments are designed to produce dose-response curves. Unfortunately, data processing approaches of the standardized toxicological tests are not uniform. There are differences between recommendations of toxicological agencies in Europe and in the United States. European OECD still supports data processing based on comparing responses each other. On the other side, American US EPA supports data evaluation of results with some algorithm based on curve fitting procedures. Situation, when experiment is evaluated only from several points which represent linear part of curve, still occurs.

This paper reviews available problem solutions and shows recommendation. Problem solution was implemented into the developed software for dose-response data processing.

PROCESSING OF DOSE RESPONSE DATA

For most types of toxic effects, it is generally considered that there is a dose or concentration below which adverse effects will not occur. Thus, the dose that can be considered as a first approximation of the theoretical threshold is critical. One such estimate of the threshold is called no observed effect level (NOEL). A NOEL for an experiment is the highest experimentally determined exposure level at which there is no statistically or biologically significant increase in the frequency or the severity of adverse effects over the control level. Some effects may be produced at this level, but they are not considered adverse or precursors to adverse effects (U.S. EPA, 1995). It is important to remember that a NOEL is an estimated value and not necessarily the same as true risk-free value. If is not possible to estimate a NOEL, a lowest observed effect level (LOEL) is estimated instead and extrapolated to a NOEL by an extrapolating factor.

The use of a NOEL approach in risk assessment has several limitations:

- The NOEL, by the definition, must be one of the experimental doses.
- If fewer animals are tested per dose group, it is less likely that real difference in response rates between two dose groups will be detected. Thus experiments with fewer animals per dose group will tend to find larger NOELs than experiments with more animals group.
- Dose spacing is another important limit. Experimental design quality strongly constraints availability and precision of the NOEL value.
- Characteristics of the dose response curve, i.e. shape and slope, are not considered in determination of the NOEL.
- Use of a NOEL does not provide estimates of the potential risk associated with any exposure level.
- NOEL approach does not make full use of the available data.

Limitations of a NOEL in risk assessment have prompted development of alternatives. Crump (1984) has suggested one of the alternatives to traditional NOEL methodology. In this approach, a dose-response or dose-effect curve is created based on experimental data and used to designate a specific dose that corresponds, for example, to a 1, 5 or 10% increased risk above control. The benchmark dose (BMD) is the statistical lower confidence limit of the dose that corresponds to a specified change in an adverse response in untreated animals (Crump, 1995). The specific change in response level is called benchmark response (BMR) and the dose that corresponds to it is called effective dose x (EDx), where x is BMR.

Since that the methodology has been modified and extended in several ways. Although there is a increasing interest in the benchmark dose approach, it has not yet found its way into “regulatory toxicology” in Europe.
(Dekkers et al., 1998), but in the United States, recent risk assessments by the U.S. EPA are made in accordance with the benchmark dose method if possible (U.S. EPA, 2003).

**BENCHMARK DOSE CALCULATION**

Benchmark dose calculation has several steps. The first step is fitting data to selected model. Thus, dose-response curve is created. We already can see several advantages of BMD approach in the first step:

- All experimental data are used for calculation, thus no information is being wasted.
- Biological principle of measured effect, such as shape and slope of dose-response curve, is taken into account in selection of the fitted model.
- Assessment of quality of final values – we can calculate confidence limit for arbitrary risk level EDx
- Simple visual comparing of the models each other – not as much dependent on experimental designed as in NOEL approach

The second step is selection of certain prespecified small response, e.g. in interval 1-10%.

The third and final step is calculation of BMD value itself. It can be realized with three different ways (Moerbeek et al., 2004):

- **Bootstrap method** – generates large number of artificial data sets by random sampling from selected probability distribution. Artificial data sets have mean dictated by fitted regression model and variance equals to residual variance of the fitted model. Each data set is based on the original experimental design, i.e. number of dose groups and numbers of subjects within dose groups have to be same. It is needed to run fitting procedure for every single generated data set. Total number of bootstrap runs that is needed to obtain stable confidence limit for any BMD value is approximately 2-10 thousand runs. Thus, bootstrap method has high time and computer hardware complexity.

- **The delta method** – calculates confidence intervals from the asymptotic covariance matrix of the estimated model parameters. Estimated confidence intervals are, by definition, symmetric around dose EDx, which do not correspond to real state.

- **Inversion method** – first constructs upper confidence limit (UCL) of the risk function for every value of dose. Next we inverse this pointwise constructed upper confidence limit to function of response and determining estimation of EDx for selected value of response.

- **The likelihood-ratio method** – searches the dose where ratio $2 \ln (L_{\text{max}}/L)$ is equal to $\chi^2_{1-\alpha,1}$, where $L_{\text{max}}$ is the maximum value of likelihood function $L$ and $\chi^2_{1-\alpha,1}$ is quantile of $\chi^2$ distribution with one degree of freedom. These confidence intervals cannot be calculated directly and some numerical method has to be used. MLE function for each of iterations of numerical method has to be determined. Time complexity of likelihood-ratio method depends on used iterative method. Despite of recursive calculation of MLE function, likelihood-ratio method is very effective and precise.

Inversion method supposes normality of sample distribution; on the other hand the likelihood-ratio method can be used for sample with distribution from family of exponential distributions.

Both Inversion method and likelihood-ratio method produce comparable results. Coverage achieved by the inversion-based BMD is at as great as coverage achieved by likelihood ratio method, thus inversion method is more conservative.

By calculating the BMD value with the likelihood-ratio method, the quality of the study is accounted for. Increasing the number of subjects in the study will result in a narrower confidence interval, and thus the higher
value of BMD. This is opposite effect of the same situation in NOEL approach, where experiments with the fewer subjects in each dose group the higher NOEL is determined.

SOFTWARE ADRIS
Software ADRIS (Assistant for Dose Response analysIS) has been developed according to U.S. EPA recommendations. Experiments can be fitted to one of the seven categorical curve shapes. Every category contains at least two models reviewed from literature or recently developed. Total number of implemented models in all categories is nineteen.

Fitting procedure proceeds according to Levenberg-Marquardt algorithm. There are two loss (penalizing) functions implemented in the software. First is common known least squares error function (LSE):

\[ LSE(\beta) = \sum_{i=1}^{n} (Y_i - \bar{Y}_i)^2, \]

where \( n \) is sample size and \( Y_i \) is the theoretical response calculated according to chosen model with vector of free parameters \( \beta \). Least square error has one important constraint – normal distribution of the sample. Natural generalization of LSE constraint is maximum likelihood function (MLE) which was implemented as the second loss function:

\[ L(\beta) = \prod_{i=1}^{n} f(Y_i | \theta), \]

where \( n \) is sample size, function \( f \) is sample distribution and \( \theta \) is vector of distribution parameters. For normal distribution general MLE function can be rewritten:

\[ L(\beta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{(Y_i - \mu(d_i))^2}{2\sigma^2} \right), \]

where \( \mu \), mean, is the value of model function in dose \( d \).

Theoretical values of LSE function are in interval \([0; +\infty)\). Quality of model fitting is criticised according the rule the smaller result value the better quality. Theoretical values of MLE function are in the interval \([0; 1]\). Quality of model fitting is criticised according the rule the higher result value the better quality. Regarding to values very close to zero, negative logarithm of MLE value is sometimes used. Then model with smaller MLE value is rewarded.

Fitted models are compared by Akaike’s information criterion (AIC) (Zucchini, 2002). AIC for MLE method is given by:

\[ AIC = -2\log L(\beta) + 2k, \]

where \( L \) is maximized likelihood function for fitted model and \( k \) is number of model free parameters. AIC for LSE method is given by:

\[ AIC = n\ln \left( \frac{LSE(\beta)}{n} \right) + 2k, \]

where \( n \) is sample size and \( LSE(\beta) \) is minimal LSE function for fitted model. According to these functions, model with the lowest AIC value is rewarded.
Benchmark dose values are calculated using likelihood ratio approach. Lately there was implemented database module in the software. Now it is possible to connect many types of database systems using ADO connection and analyse raw data directly from database system.

Software ADRIS is written in C++ programming language.

ACKNOWLEDGEMENTS

The study was supported by the Ministry of Education of the Czech Republic (project No. MSMT 0021622412 INCHEMBIOL).

REFERENCES


INFORMATION SOURCES FOR EFFECT ASSESSMENT

RISK ASSESSMNET
http://riskassessment.ornl.gov/hhra.cfm

Ecological Risk Analysis – Guidance, tools, and Applications

ChemFinder: http://chemfinder.cambridgesoft.com/

Chemical-Specific Toxicity Values (The Risk Assessment Information System)
http://risk.lsd.ornl.gov/tox/tox_values.shtml

Compendium of Pesticide Common Names - Index of Common Names
http://www.hcrs.demon.co.uk/index_cn_frame.html
http://www.hcrs.demon.co.uk/index.html

IRIS – database U.S. EPA
http://www.epa.gov/iris/

IRIS – U.S. EPA - List of available Toxicological Reviews
http://www.epa.gov/iris/toxreviews/index.html

Ecotoxicological database „PLUMBUM“ (in czech)

ASTDR - Agency for Toxic Substances and Disease Registry
http://www.atsdr.cdc.gov/toxfaq.html

ITER - TERA - Toxicology Excellence for Risk Assessment
http://www.tera.org/

ECDIN - Environmental Chemicals Data and Information Network
http://ecdin.etomep.net/

International Hazard Datasheets on Occupations
MRL – Minimal Risk Levels (ASTDR)

NASD – databáze CD (National Agriculture Safety Database)
http://www.cdc.gov/niosh/nasd.html

IDLHs – (Immediately dangerous to life or health)
http://www.cdc.gov/niosh/idlh/idlh-1.html

NIOSH - ICSC (International Chemical Safety Cards)
http://www.cdc.gov/niosh/ipcs/ipcs0000.html
http://www.cdc.gov/niosh/ipcs/icstart.html

Pesticide databases – „Recognition and Management of Pesticide Poisonings“
http://ace.orst.edu/info/nptn/rmpp.htm

ExToxNET - The EXTension TOXicology NETwork
http://ace.orst.edu/info/extoxnet/

TOXNET – Toxicology Data Network - Databases on toxicology, hazardous chemicals, environmental health, and toxic releases

NIOSH - National Institute for Occupational Safety and Health
http://www.cdc.gov/niosh/vendors.html

ACGIH – ACGIH - American Conference of Governmental Industrial Hygienists
http://www.acgih.org/free/welcome.htm

SIS-Specialized Information Services Division
http://sis.nlm.nih.gov/

ILO-CIS: International Occupational Safety and Health Information Centre

REPROTOX® - An Information System on Environmental Hazards to Human Reproduction and Development
http://www.reprotox.org/

Genotoxicity Database (Ames test, SOS – Chromotest)
http://www.pasteur.fr/recherche/unites/pmtg/toxic/database.iphtml

Gen-Tox: Genetic Toxicology database

IARC (International Agency for Research on Cancer)
   Cancer Databases and other Resources
http://www.iarc.fr/

CAS – Chemical Abstract Service
http://info.cas.org/
ENCLOSURE: A GUIDE TO DOSE-RESPONSE MODELLING FOR ECORA USERS IN R

BRIEF THEORY SUMMARY
The theoretical introduction is based on “The Bioassay Analysis using R and Bioassay for Allelochemicals: Examples with R” written by Christian Ritz and Jens C. Streibig.

R (www.r-project.org) is a language and environment for statistical computing and graphics. It is a GNU project which is similar to the S language and environment which was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and colleagues. R can be considered as a different implementation of S. There are some important differences, but much code written for S runs unaltered under R.

R provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering, ...) and graphical techniques, and is highly extensible. The S language is often the vehicle of choice for research in statistical methodology, and R provides an Open Source route to participation in that activity.

One of R's strengths is the ease with which well-designed publication-quality plots can be produced, including mathematical symbols and formulae where needed. Great care has been taken over the defaults for the minor design choices in graphics, but the user retains full control.

R is available as Free Software under the terms of the Free Software Foundation's GNU General Public License in source code form. It compiles and runs on a wide variety of UNIX platforms and similar systems (including FreeBSD and Linux), Windows and MacOS.

Dose-response curve: a model of a relationship between the dose (several doses of one chemical, fertilizer, etc.) and the response – studied reaction (e.g. percentage of dead cells, intensity of activity)

Example of dose-response curve: The logistic dose-response curve a) on non-logarithmic dose scale; b) on logarithmic dose scale.

The quantities effective dosage (ED) and selectivity index (SI) are commonly used to compare different chemicals. EDy is defined as the dose that yields a response which is (100-y)% of the maximal response (a reduction of y%). SI(x,y) is the ratio between EDx for one curve and EDy for another curves.

\[ SI(x,y) = \frac{EDx}{EDy} \]

Example: Vertical a) and horizontal b) comparison of dose-response curves for two biological active compounds in a plant species.
USED MODELS SUMMARY

The three-parameter exponential function

\[ f(x, (b_0, b_1, b_2)) = b_0 + \exp(b_1 + b_2 x) \]

The three-parameter logistic function

\[ f(x, (\text{asym}, \text{xmid}, \text{scal})) = \frac{\text{asym}}{1 + \exp \left( \frac{\text{xmid} - x}{\text{scal}} \right)} \]

The four-parameter logistic function is given by the formula with parameters b,c,d,e

\[ f(x, (b, c, d, e)) = c + \frac{d - c}{1 + \exp \left( b \left( \log(x) - \log(e) \right) \right)} \]

Brain-Cousens' model is used in situations where hormesis is present:

\[ f(x, (b, c, d, e)) = c + \frac{d + f x - c}{1 + \exp \left( b \left( \log(x) - \log(e) \right) \right)} \]

Brain-Cousens model is obtained by modifying the four-parameter logistic model by adding the linear term fx in the numerator.

NOTATION

This executable R code is available in file Dose-response curves&R_v7.r, the version of R needed is at least R 2.3.0 or higher version with library drc downloaded (www.r-project.org), the example data are in the datafile Dataset_Zlin_b.csv.

The screenshot of the code are given in red color, the results provided by R are given in black color and the additional comments on the results are given in green color.

Two more notes on R code:

- # is the comment sign
- > is the sign for a line of R code, lines without this sign contain the results of computation.
DOSE-RESPONSE ANALYSIS IN R: EXAMPLE

STEP 1: DATA READING

```r
> # R is a free, object oriented, statistical software
> # Dowload: http://cran.r-project.org/
> #
> #-------------------Preparation for our work--------------------------------------
> #
> # Set working directory
> #setwd("C:/Dose-response_R")
> #
> # Read data in a text file
> data<-read.table("Dataset_Zlin_b.csv",sep="",header=TRUE)
> data

Metabolic_activation Season Locality Dose Response
1 without S9 Summer Vsetin 15.000 2.00
2 without S9 Summer Vsetin 15.000 1.81
3 7.500 1.53
......
412 with S9 NA NA
```

STEP 2: DATA SELECTION

```r
> # Select localities of interest
> attach(data)
> Vsetin_summer1<-subset(data,(Locality=="Vsetin")&(Season=="Summer")&(Metabolic_activation=="without S9"), select=c(Dose, Response))
> Vsetin_summer1

Dose Response
1 15.000 2.00
2 15.000 1.81
3 15.000 2.06
4 7.500 1.53
......

>Bynina_summer1<-subset(data, (Locality=="Bynina")&(Season=="Summer")&(Metabolic_activation=="without S9"), select=c(Dose, Response))
> Bynina_summer1
```
Mstenovice_summer1<-subset(data, (Locality=="Mstenovice")& (Season=="Summer")& (Metabolic_activation=="without S9"), select=c(Dose, Response))
Mstenovice_summer1
detach(data)

STEP 3: BASIC DATA VISUALISATION AND INSPECTION

# data visualisation
plot(Response~Dose,type="p",data=Vsetin_summer1,xlab="Dose",ylab="Response",main="Vsetin - summer - without S9")

# Set the History - recording
plot(Response~log(Dose),type="p",data=Vsetin_summer1,xlab="log(Dose)",ylab="Response",main="Vsetin - summer - without S9")

# Log transformation on x-axis
The log-transformation on x-axis is commonly used, when the points of measurement are not equidistant.
This chart depicts the non-homogeneous variance of measurements.

STEP 4: FITTING NON-LINEAR REGRESSION (DOSE-RESPONSE CURVE) IN R

> # Dose-response curve = a (special) type of nonlinear regression with normal distribution error
> # Standart fitting of nonlinear regression models in R: function nls
> help(nls) # help on the command nls
>
> ############Data Mstenovice###################################
> # Exponencial model with 3 parameters: b0, b1, b2
> # Gauss-Newton algorithm
> # starting values of iterative algorithm: 1,1,1
> model_e3 <- nls((Response) ~ b0 + exp(b1 + b2 * log(Dose)), data = Mstenovice_summer1, start = list(b0 = 1, b1 = 1, b2 = 1))
> summary(model_e3)

Formula: (Response) ~ b0 + exp(b1 + b2 * log(Dose))

Parameters:

| Estimate | Std. Error | t value | Pr(>|t|) |
|----------|------------|---------|----------|
| b0 1.2170 | 0.3020     | 4.030   | 0.00297 ** |
| b1 -1.7615 | 1.2028   | -1.464  | 0.17710 |
| b2 0.6880  | 0.3527    | 1.951   | 0.08286 . |

The results of model show that the exponential trend is non-significant.

---

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.108 on 9 degrees of freedom

> #summary(model_e3)$residuals

> plot(summary(model_e3)$residuals ~ log(Dose), data = Mstenovice_summer1, type = "p", xlab = "log(Dose)", ylab = "Residuals", main = "Mstenovice - summer - without S9")

The residual analysis is important in any regression analysis, in this case residuals do not provide any trend and shows random fluctuation around 0; it means correct model was applied.

> plot((Response) ~ log(Dose), data = Mstenovice_summer1, type = "p", xlab = "log(Dose)", ylab = "Response", main = "Mstenovice - summer - without S9")

> minld <- log(1.875)
> maxld <- log(15)
> xx <- seq(minld, maxld, by = 0.01)
> b0 <- summary(model_e3)$coefficients[1, 1]
> b1 <- summary(model_e3)$coefficients[2, 1]
> b2<-summary(model_e3)$coefficients[3,1]
> lines(b0+exp(b1+b2*xx)^xx,col="red",type="l")
> Response_of_interest<-1.5  Response 1.5 is defined to be important.
> Dose_of_interest<-((log(Response_of_interest-b0)-b1)/b2
> # on LOG scale!!!!
> Dose_of_interest
> [1] 0.7254417  estimated dose for response 1.5
> lines(rep(Response_of_interest,length(xx))^xx,col="grey", type="l")
> pom<-seq(0,Response_of_interest,by=0.1)
> lines(pom^rep(Dose_of_interest,length(pom)),col="grey", type="l")

---

> # Logistic model model with 3 parameters - pre-defined function> Sslogis
> plot((Response)~log(Dose),data=Bynina_summer1, type="p",xlab="log(Dose)",ylab="Response",main="Bynina - summer - without S9")
> model_l3 <- nls(Response ~ SSlogis(log(Dose), Asym, xmid, scal), data=Bynina_summer1)
> summary(model_l3)

Formula: Response ~ SSlogis(log(Dose), Asym, xmid, scal)

Parameters:

| Estimate | Std. Error | t value | Pr(>|t|) |
|----------|------------|---------|----------|
| Asym     | 33.971     | 1503.056| 0.023    | 0.982    |
| xmid     | 14.758     | 231.766 | 0.064    | 0.951    |
| scal     | 4.354      | 9.814   | 0.444    | 0.668    |

Residual standard error: 0.1472 on 9 degrees of freedom

The results of model show that the logistic trend is non-significant.
> mind<-1.875
> maxd<-15

#plot(summary(model_l3)$residuals~log(Dose),data=Bynina_summer1,type="p",xlab="log(Dose)",ylab="Residuals",main="Bynina - summer - without S9")

> # A problem - large variability of responses for dose = 15
> xx<-seq(minld,maxld, by=0.01)
> Asym<-summary(model_l3)$coefficients[1,1]
> xmid<-summary(model_l3)$coefficients[2,1]
> scal<-summary(model_l3)$coefficients[3,1]
> lines(Asym/(1+exp((xmid-xx)/scal))~xx, type="l",col="red")

> Response_of_interest<-1.5
> Dose_of_interest<- -scal*(log(Asym/Response_of_interest-1))+xmid

[1] 1.370003 Estimated dose for response 1.5

> # on LOG scale!!!!
> lines(rep(1.5,length(xx))~xx,col="grey", type="l")
> pom<-seq(0,Response_of_interest,by=0.1)
> lines(pom~rep(Dose_of_interest,length(pom)),col="grey", type="l")

> # Exponencial model with 3 parameters: b0, b1, b2
>
> # Gauss-Newton algorithm
>
> # starting values of iterative algorithm: 1,1,1
> model_e3_B<-nls((Response)~b0+exp(b1+b2*log(Dose)), data= Bynina_summer1 , start=list(b0=1,b1=1,b2=1))
> summary(model_e3_B)

Formula: (Response) ~ b0 + exp(b1 + b2 * log(Dose))

Parameters:
Estimate Std. Error t value Pr(>|t|)
\[b_0\] 0.13406 3.15059 0.043 0.967
\[b_1\] -0.01625 3.03279 -0.005 0.996
\[b_2\] 0.23830 0.50774 0.469 0.650
Residual standard error: 0.1472 on 9 degrees of freedom

The results of model show that the exponential trend is non-significant.

```r
> #summary(model_e3_B)$residuals
> #plot(summary(model_e3_B)$residuals~log(Dose),data=Bynina_summer1,type="p",xlab="log(Dose)",ylab="Residuals",main="Bynina - summer - without S9")
> plot((Response)~log(Dose),data=Bynina_summer1,type="p",xlab="log(Dose)",ylab="Response",main="Bynina - summer- without S9")
> minld<-log(1.875)
> maxld<-log(15)
> xx<-seq(minld,maxld, by=0.01)
> b0<-summary(model_e3_B)$coefficients[1,1]
> b1<-summary(model_e3_B)$coefficients[2,1]
> b2<-summary(model_e3_B)$coefficients[3,1]
> lines(b0+exp(b1+b2*xx)~xx,col="red",type="l")
> lines(Asym/(1+exp((xmid-xx)/scal))~xx, type="l",col="green")
> Response_of_interest<-1.5
> Dose_of_interest<-(log(Response_of_interest-b0)-b1)/b2
> Dose_of_interest
[1] 1.376845 Estimated dose for response 1.5
> # on LOG scale!!!!
> lines(rep(Response_of_interest,length(xx))~xx,col="grey", type="l")
> pom<-seq(0,Response_of_interest,by=0.1)
> lines(pom~rep(Dose_of_interest,length(pom)),col="grey", type="l")
```
Bynina - summer - without 89

> # Linear regression model
> model_lm<-lm(Response~log(Dose),data= Bynina_summer1)
> summary(model_lm)

Call:
  lm(formula = Response ~ log(Dose), data = Bynina_summer1)

Residuals:
   Min     1Q  Median     3Q    Max
-0.249333 -0.022333 -0.003333  0.054167  0.280667

Coefficients:
                     Estimate Std. Error t value Pr(> |t|)
(Intercept)      1.03605     0.09706  10.675 8.71e-07 ***
log(Dose)           0.35202     0.05276   6.672 5.56e-05 ***

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Dose is significant independent variable

Residual standard error: 0.1416 on 10 degrees of freedom
Multiple R-Squared: 0.8166,  Adjusted R-squared: 0.7982
F-statistic: 44.51 on 1 and 10 DF,  p-value: 5.556e-05

> a<-model_lm$coefficients[1]
> b<-model_lm$coefficients[2]
> lines(a+b*xx~xx, type="l",col="black" )
STEP 5: SPECIALIZED LIBRARIES FOR ANALYSIS OF DOSE-RESPONSE CURVES IN R

> #-------------------Libraries for dose-response curves------------------------------------------
> # Libraries in R specialised on analysis of dose-response curves
> library(drc)    # Analysis of dose-response curves
>                   # http://www.bioassay.dk/
>                   # R version 2.3.0
> Loading required package: MASS

> # in progress
> library(drfit)  # Dose-response data evaluation
>                   # http://cran.r-project.org/
> Loading required package: RODBC

> # also usefull
> library(multcomp) # The multcomp package allows for multiple comparisons of k groups in general linear models.
>                   # angina - dose response data
> Loading required package: mvtnorm
> #----------------------------Library drc---------------------------------------------
> # Read library
> library(drc)
> ################# Read data FA #################
> data(FA)
> FA
> MEANLR   MM
> 1   7.5800000 0.00
> ..... 
> help(FA)
> plot(MEANLR~MM,data=FA,xlab="Dose",ylab="Response",main="FA data")

> # Fitting a 4-parameter logistic model
> ## with Box-Cox transformation
> model_FA_l4 <- multdrc(FA, boxcox = TRUE)
> summary(model_FA_l4)
A 'logistic' model was fitted.
Parameter estimates:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>b:(Intercept)</td>
<td>2.499575</td>
<td>0.373285</td>
<td>6.696153</td>
<td>1.620e-06</td>
</tr>
<tr>
<td>c:(Intercept)</td>
<td>0.366576</td>
<td>0.094803</td>
<td>3.866715</td>
<td>0.001</td>
</tr>
<tr>
<td>d:(Intercept)</td>
<td>7.900308</td>
<td>0.338958</td>
<td>23.307638</td>
<td>6.184e-16</td>
</tr>
<tr>
<td>e:(Intercept)</td>
<td>2.989462</td>
<td>0.229544</td>
<td>13.023480</td>
<td>3.159e-11</td>
</tr>
</tbody>
</table>

Estimated residual variance: 0.07634711
Heterogeneity adjustment: Box-Cox transformation

Estimated lambda: 0.4
Confidence interval for lambda: [0.126, 0.782]

> plot(model_FA_l4)

> ED(model_FA_l4,c(10,50,90))  # estimate of effective dose – 10%, 50%, 90% of maximal response

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:10</td>
<td>1.2412</td>
<td>0.2221</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1:50</td>
<td>2.9895</td>
<td>0.2295</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1:90</td>
<td>7.2004</td>
<td>0.8592</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

> # Brain-Cousens modification of logistic function  # hormesis effect
> model_FA_BC <- multdrc(FA, fct=braincousens(fixed=c(NA, NA, NA, NA, NA),lowerc=c(1, -Inf, -Inf, -Inf, -Inf)), control=mdControl(constr=TRUE))
> summary(model_FA_BC)

A 'braincousens' model was fitted.

Parameter estimates:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>b: (Intercept)</td>
<td>2.75686</td>
<td>0.52461</td>
<td>5.25510</td>
<td>4.511e-05</td>
</tr>
<tr>
<td>c: (Intercept)</td>
<td>0.41515</td>
<td>0.26309</td>
<td>1.57794</td>
<td>0.1311</td>
</tr>
<tr>
<td>d: (Intercept)</td>
<td>7.74173</td>
<td>0.21415</td>
<td>36.15085</td>
<td>2.773e-19</td>
</tr>
<tr>
<td>e: (Intercept)</td>
<td>2.77684</td>
<td>0.50940</td>
<td>5.45123</td>
<td>2.931e-19</td>
</tr>
<tr>
<td>f: (Intercept)</td>
<td>0.35676</td>
<td>0.68830</td>
<td>0.51832</td>
<td>0.6102</td>
</tr>
</tbody>
</table>

Estimated residual variance: 0.2795121

> plot(model_FA_BC)
The choice of model is crucial - compare the estimates of EDs with 4-parameter logistic model.

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1:10</td>
<td>1.5318</td>
<td>0.2021</td>
<td></td>
</tr>
<tr>
<td>1:50</td>
<td>3.0517</td>
<td>0.1951</td>
<td></td>
</tr>
<tr>
<td>1:90</td>
<td>6.9143</td>
<td>1.4758</td>
<td></td>
</tr>
</tbody>
</table>

Detailed explanation of applied algorithms and other extensive examples are available in papers “The Bioassay Analysis using R and Bioassay for Allelochemicals: Examples with R” written by Christian Ritz and Jens C. Streibig; the papers are available through www.r-project.org.
IV. RISK CHARACTERIZATION

The “risk characterization” is the third major part of the EcoRA process in which we can utilize Geographic Information System. The risk characterization is the final stage of the ecological risk assessment and it is a conclusion made from all the previous steps. Risk characterization allows risk assessors to determine the relationships between stressors, effects and ecological entities and to reach conclusions regarding the occurrence of exposure and the adversity of existing or anticipated effects. Once the risk characterization is finished, we should be able to estimate the ecological risk and interpret the adversity of ecological effects.

GIS is the right instrument to elaborate and visualize these outputs. A simple display of risk for different receptors and contaminants in different times is shown in presented subsequent maps and outputs.

Figure 1a) Risk maps of the selected POPs contamination with regard to soil limits.
Our study showed the impact of the anthropogenic activities on the environmental contamination. We have monitored contamination levels in sediments and alluvial soils across the Zlín region. Comparison to the national soil limits shows that soils from multiple sites repeatedly contained greater contamination (especially
in the industrial area) than the background limits for agricultural soils, namely for PAHs and PCBs. The concentrations in sediments are also high.

We focused on river system, where hazardous hydrophobic contaminants are predominantly transported in association with suspended particulate matter that accumulates in regions of low turbulence. The accumulated material can be greatly redistributed during erosion events. The floods in 1997 with their extreme extent had really significant effects on the contamination level. The studied soil sites were not affected by any other floods before or after the event investigated in 1997 during the sampling period.

The floods significantly affected the distribution of pollutants. On the risk maps we can see the differences before and after the floods. PCBs were imprimis washed out, while PAHs and DDTs have moved within the region.

Figure 2a) Spatial and time related trend of real contamination of alluvial soils with PAHs – hot spots.
Figure 2b) Spatial and time related trend of real contamination of alluvial soils with PCBs – hot spots.

This maps show the spatial and time related trend of real contamination of alluvial soils with PAHs and PCBs in comparison to the national limits. The most contaminated area of the Zlín – Otrokovice with hot spot sources Barum Continental and Hamag is displayed. The contamination was several times greater than the background limits (especially in 1997). The impact of the floods in July 1997 is evident (especially the PCBs washing out).
Figure 3) Hazard quotients for some of the selected receptors in the most contaminated localities

Hazard quotients for some PAHs (Anthracen, Benzo-a-pyrene and Fluoranthen) and receptors (Poaceae, Lumbricidae, soil species) are shown on these exemple maps. The values for the “hot spot” sampling sites before and after the floods are displayed.

The contamination with Anthracen is on the level that doesn’t represent a risk for the growth of the species from the family Poaceae (except one locality down the river from Hamag after the floods).

At this locality the highest hazard quotients for other PAHs and receptors were also find out. The washing out of the industrial source (Hamag) and remobilization of contaminated sediments down the river is evident.
In case of the Benzo-a-pyrene, risk for survival and reproduction of *Lumbricidae* was calculated. In the most cases the risk was significant. The influence of floods is apparent – remobilisation and deposition is dependent on the contaminant source distance.

In respect to the risk values for Fluoranthen, the effects on all soil species and the effects on the mortality of *Lumbricidae* were determined. The risk limit has been exceeded almost at each locality (especially for „soil species“). Of course HQs for „species“ are in comparison to HQs for *Lumbricidae* sufficiently safe to protect more sensitive species.

**CONCLUSIONS**

Presented results and the other outcomes showed the consequential contamination of study region with POPs. The sources are transportation, solid fuels combustion, agriculture but firstly the industrial establishments. Except the contamination level compared to the valid limits, the risk quotients for selected receptors (endpoints) and substances were determined. The risk values depend on the substances, regarding to the urban agglomeration and the hot-spot sources the counted risk values almost in all cases exceeded the risk limit. The soil concentration of the monitored pollutants potentially constitute a hi-risk for soil species and soil processes, therefore for the soil quality and affected ecosystems too and this problem should be timely solved.

In addition, the region was affected by the floods that significantly influence the distribution of the pollutants. The floods cause a remobilization of contaminated sediments. Those are subsequently deposited on the inundated areas and it probably results in an increase in soil contamination.