

INFORMAČNÍ SYSTÉM HODNOCENÍ EKOLOGICKÉHO STAVU POVRCHOVÝCH VOD

Karel Kiswa¹, Jaroslav Ráček²

¹Fakulta informatiky, Masarykova univerzita

²Institut biostatistiky a analýz, Masarykova univerzita

ABSTRAKT

Projekt ARROW (Assessment and Reference Reports of Water monitoring) je systémovým řešením pro sledování a hodnocení ekologického stavu vod. Odráží v sobě požadavky WFD a současně i poptávku po komplexním řešení problematiky od sledování jednotlivých biologických a hydromorfologických složek až po sběr a hodnocení dat. Nedílnou součástí je také kalibrace a nastavení referenčních modelů pro hodnocení. Projekt navazuje na dlouhodobý vývoj problematiky v České republice a je současně i výsledkem paralelních projektů řešených pod záštitou Ministerstva životního prostředí v rámci přípravy programů monitoringu. V tomto příspěvku je prezentována metodologie, technologie a způsob vývoje výpočetní knihovny, která pomáhá řešit jak kalibraci a hodnocení referenčních lokalit tak také hodnocení lokalit nereferenčních.

ABSTRACT

The ARROW (Assessment and Reference Reports of Water monitoring) project is the project of implementation of WFD in the monitoring of surface waters of the Czech Republic and covers all aspects of this problem from data sampling to informatics solution (focusing on the ecostat). The project incorporates both WFD demands and complex solution of sampling, biological, analytical and ICT problems. The integral part of the project is also the definition and calibration of reference network of sites. The project is based on long term development of this field in the Czech Republic and is related to parallel project of Ministry of the Environment of the Czech Republic connected to monitoring programs. Methodology, technologies and the way of development of the Computational library which helps to solve calibration and evaluation of reference sites as well as evaluation of new (non-reference) sites are being discussed in this paper.

KLÍČOVÁ SLOVA:

ARROW, XML, Java, hodnocení stavu, kalibrace, hodnotící metriky.

KEY WORDS

ARROW, XML, Java, evaluation, calibration, evaluation metrics.

1. ÚVOD

Česká republika byla povinna v roce 2006 zavést systém sledování a hodnocení ekologického stavu povrchových vod v souladu s požadavky jednotné implementační strategie Rámcové směrnice vodní politiky Evropských společenství (2000/60/EC).

S touto povinností souvisí i potřeba kvalitní metodiky vyhodnocení dat monitoringu a biomonitoringu sledovaných za tímto účelem a tedy faktické naplnění požadovaných výstupů hodnocení kvality povrchových vod.

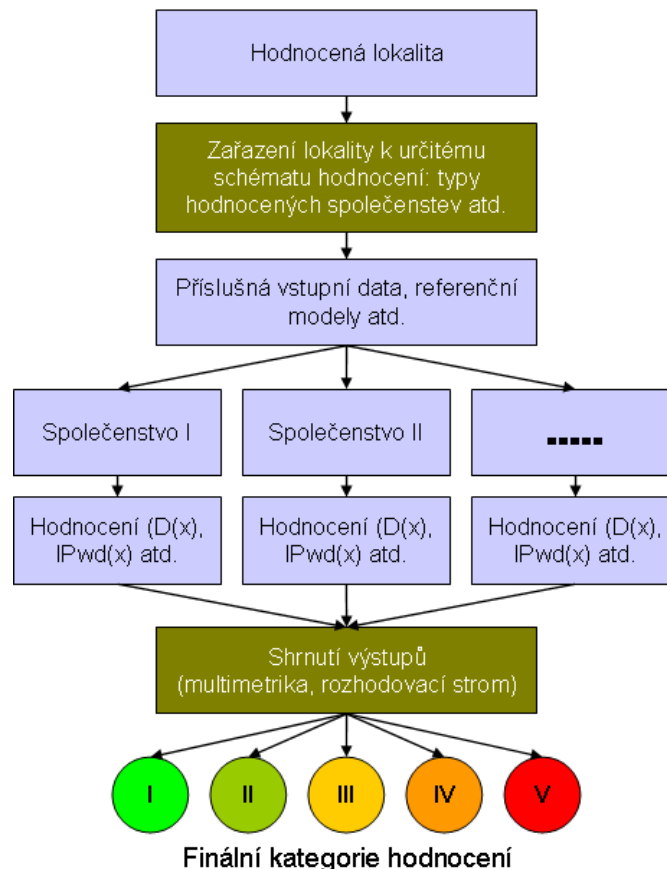
Problém je komplikován faktem, že implementovaný systém hodnocení musí být schopen hodnotit data sbíraná v souladu s požadavky jednotné implementační strategie Rámcové směrnice vodní politiky Evropských společenství (2000/60/EC).

Je téměř jasné, že spojení takto různorodých pohledů na hodnocení stavu povrchových vod bude vyžadovat multimetrický přístup kombinující data z biologických, hydromorfologických a chemicko-fyzikálních pohledů, přičemž každý z těchto pohledů ponese svoji vlastní metodiku hodnocení. Jako nejperspektivnější metodou hodnocení jednotlivých společenstev se jeví srovnání lokalit biomonitorovací sítě s referenčními („přirozenými“) daty a určení míry narušení přirozeného stavu lokalit srovnáním skutečného a očekávaného („přirozeného“) stavu.

2. OBECNÉ SCHÉMA HODNOCENÍ

Celý průběh hodnocení lze považovat za velmi komplexní proces, kde vlastní matematické metody tvoří pouze menší část celého problému. Vlastnímu zahájení analýzy musí předcházet definice „schémat hodnocení lokalit“, které respektuje vhodnost nebo vůbec využitelnost různých typů společenstev pro různé typy toků.

Teprve po této fázi, kdy je pro hodnocenou lokalitu určena sada biologických společenstev a dalších dat, je možné přistoupit k vlastní analýze dat. Analýza probíhá pro různá biologická společenstva a typy dat odděleně a po jejím skončení jsou dílčí výsledky různých částí analýzy shrnuty do formy jednoduchého kategoriálního hodnocení stavu (viz. následující obrázek).



Obr. 1: Obecný model procesu hodnocení ekologického stavu vod.

3. TECHNICKÁ REALIZACE

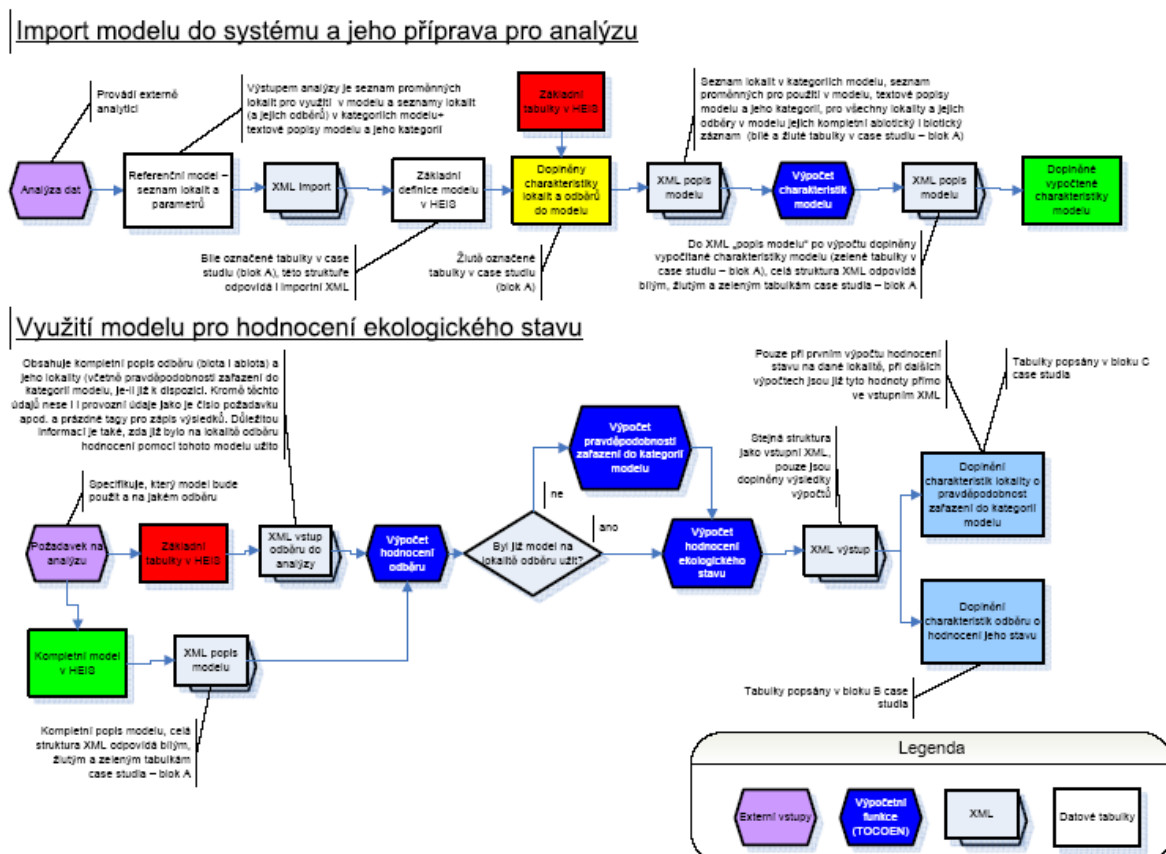
Řešitelský tým navrhnul vyvíjet hodnotící systém formou samostatné výpočetní knihovny. Hodnotící systém je implementován v jazyce Java a s databází bude komunikovat skrze standardizované XML rozhraní, které je v podstatě obrazem databázových tabulek, které obsahují data nezbytná pro výpočet. Tímto řešením datové základny odpadá problém se synchronizací zrcadlené databáze a tím i případnou nekonzistencí dat ze strany vývoje.

Během testovacího provozu se plánuje provoz knihovny jako webové služby (využívající protokol HTTP) poskytované na straně vývoje a využívané zadavatelem pomocí webového klienta. Po skončení projektu se knihovna bude implementovat na straně zadavatele, kde se bude spouštět lokálně a podle potřeby si zadavatel implementuje vlastní webové služby.

Díky standardizovanému rozhraní mohou s knihovnou v budoucnosti komunikovat prostřednictvím těchto webových služeb i další subjekty, které ji mohou využívat i pro hodnocení dalších lokalit, které nejsou zaznamenány v databázi centrálního IS.

Provedené výpočty budou zasílány zpět centrálnímu IS opět formou XML souborů. Veškeré vstupní i výstupní XML soubory budou zároveň archivovány, aby mohly být v budoucnu použity jako datové podklady pro tvorbu reportů.

Následující obrázek popisuje procesy, které probíhají během výpočtu hodnocení. V první části se jedná o výpočet modelu a ve druhé jde o hodnocení ekologického stavu lokality.



Obr. 2: Procesy výpočtů podporovaných hodnotící knihovnou ARROW.

4. METODY HODNOCENÍ DAT BIOLOGICKÝCH SPOLEČENSTEV

Pro analýzu vlivu prostředí a pro popis struktury a chování biologických společenstev vznikla celá řada analytických metod sahající od jednoduchých rovnic (indexy) až po komplikované metody statistické analýzy (vícerozměrná analýza dat) vyvíjených z důvodů od čistě vědeckých až po empiricky odvozené a přímo prakticky použitelné výpočty.

Pro praktické nasazení biologických společenstev při hodnocení stavu prostředí je třeba tyto přístupy ovšem vyhodnotit co do:

- Informační hodnoty, kterou přinášejí pro hodnocení stavu povrchových vod
- Uchopitelnost a interpretovatelnost výstupů
- Náročnost na podkladová data
- Výpočetní a implementační náročnost

Lze říci, že celkové řešení hodnocení stavu bude využívat prvků ze všech dvou základních přístupů:

- Indexové hodnocení
- Vícerozměrné hodnocení

Indexové hodnocení

Nejjednodušším přístupem hodnocení je shrnutí biologických společenstev pomocí jednoduchých rovnic do jednoho čísla s určitou biologickou nebo praktickou interpretací. Indexové hodnocení je možno rozdělit na indexy:

- Popisující počet jedinců a/nebo druhů ve společenstvu bez jakékoliv vazby na vlastnosti konkrétních druhů
- Kombinující výskyt nebo abundance druhů s jejich detailnějšími charakteristikami (potravní nebo habitatové preference, saprobní valence atd.)

Mezi indexy první skupiny lze zařadit v podstatě pouze indexy diverzity. Tyto indexy postihují různé aspekty struktury biologických společenstev (počet druhů, poměry početností druhů) a odráží strukturální změny společenstva. Jejichž využitelnost pro komplexní hodnotící systém je spíše pomocná protože jde o poměrně málo citlivé ukazatele nespécifických změn ve společenstvu nerespektující výskyt konkrétních druhů. V metodice hodnocení stavu povrchových vod je počítáno s výpočtem následujících základních indexů diverzity, výpočet je univerzální pro všechny typy biologických společenstev:

- Počet druhů ve společenstvu
- Shannon index a Shannon evenness
- Simpson index
- Berger-Parker index
- Stochastické Q statistics

Vícerozměrné hodnocení

Nejkomplexnější přístup analýzy biologických společenstev využívá nikoliv jednoduchých číselných shrnutí (indexů) za celé společenstvo (tedy převedení na problém jednorozměrné statistiky), ale vícerozměrné analýzy společenstev, kdy každý taxon nebo abiotický parametr představuje jeden rozměr popisující hodnocenou lokalitu. Zároveň nejsou jednotlivé lokality/společenstva hodnoceny odděleně, ale vždy ve vzájemném stavu s ostatními lokalitami/společenstvy (jde vlastně o jejich vzájemnou pozici ve vícerozměrném prostoru) a můžeme tak přesně určit v čem se hodnocená lokalita liší od ostatních (např. znečištěná od neznečištěných). Výhodou přístupu je také využití veškeré variability nasbíraných dat a možnost predikcí na úrovni jednotlivých hodnocených lokalit.

Problematika nasazení XML

Při vývoji aplikace byla třeba rozhodnout, které technologie použít. Nasazení XML jako formátu pro přenos dat bylo odsouhlaseno všemi zúčastněnými již v počátečních fázích vývoje. Hlavním důvodem pro volbu XML byla kvalitní podpora ze strany vývojového prostředí (Java) a také fakt, že pomocí XML lze velmi jednoduše sjednotit strukturu dat. Dalším silným argumentem pro použití XML byla jeho schopnost omezit variabilitu vstupních dat a tím také výrazně zjednodušit složitost následujícího výpočtu hodnotících metrik a modelu.

Dalším faktorem pro volbu XML formátu bylo vytvoření rozhraní aplikace tak, aby se vývoj modulu mohl rozdělit do dvou částí. První část byla realizována vývojovým týmem IBA, jedná se o vývoj výpočetní knihovny. Druhá část je realizována bez naší podpory (obsahem druhého modulu je svázání hodnocených dat se strukturou databáze a správa výsledků spočítaných pomocí výpočetní knihovny). Data přenášená pomocí XML posloužila jako spojující prvek naší spolupráce a pomohla sjednotit a usměrnit vývoj modulu.

V pokročilejších fázích projektu začala velikost zdrojových XML dokumentů výrazně narůstat. (z velké části byl tento jev zapříčiněn používáním dat, která v původním návrhu nebyla respektována a uvažována).

Z tohoto důvodu musel být celkově přehodnocen přístup ke zpracování takto velkých souborů. Vzhledem k času potřebnému ke zpracování jednoho dotazu, bylo použití DOM pro vyhledávání v XML již nepraktické nebo spíše nepoužitelné. Jedinou použitelnou alternativou bylo nasazení technologie SAX. Jak se ukázalo, toto rozhraní pro zpracování XML je nesrovnatelně rychlejší než DOM, což nám umožnilo zpracovávat rozsáhle soubory v řádu několika málo sekund.

Nečistota dat

I přes masivní nasazení XML a jeho restriktivních opatření nebylo možné zcela zamezit výskytu chyb ve vstupních datech. Nejčastějšími závadami se překvapivě stávala data, která ve zdrojovém XML souboru měla být obsažena (byla nutná k výpočtu metrik), ale v databázi (která sloužila jako zdroj informací) tato data nebyla uložena. Tyto nedostatky způsobovaly značné problémy, kdy například k vybranému taxonu nebylo možné dohledat některé z jeho vlastností. Tento problém byl ještě umocněn skutečností, že vlastnosti taxonů jsou organizovány do skupin v závislosti na jejich typu. To znamenalo, že v případě, kdy v některé skupině chyběla byť jenom jedna vlastnost, musely se i ostatní vlastnosti z této skupiny ignorovat.

V případech, kdy nepomohla aplikace XML Schema pro jasnou a jednoznačnou definici dat, bylo nutné řešit problém programově. Pro každou metriku, u které mohla nastat situace s nekompletními daty, se provedlo předzpracování počítaných dat a na data byly aplikovány filtry, které tato neúplná nebo chybná data odstranily. V případě, kdy byla data správná, proběhl výpočet standardním způsobem. Z důvodu objektivity spočítaných metrik bylo nezbytné, aby se pro každou funkci zaznamenávalo, kolik taxonů bylo při jejím výpočtu vyřazeno (pro každou funkci mohl být tento počet zcela jiný, protože v každé funkci se zpracovávají jiné vlastnosti taxonů).

Rychlost výpočtu

Pro podpoření rychlosti zpracování dat při výpočtu bylo nezbytné optimalizovat zdrojové kódy. Nejvýraznější prodlevy nastávaly při práci s XML. Naše práce se tedy zaměřila především na metody pracujícími se zdrojovými XML soubory. Nejvýraznějšího zrychlení jsme dosáhli nasazením technologie SAX místo dříve používaného DOM (zkrácení výpočtu z 15 minut na 30 vteřin). Ani toto pro nás nebyl uspokojivý výsledek a tak jsme výrazným způsobem optimalizovali ukládání zpracovávaných dat (aby se stejná data nemusela opakovaně načítat). Jednalo se ve své podstatě o jednoduchou formu cachování. Tímto

prostředkem byla rychlost snížena přibližně na 5 vteřin. To je již dostatečně nízká hodnota a z tohoto důvodu již další optimalizace kódu nebyla potřeba.

5. ZÁVĚR

Řešení projektu probíhá ve spolupráci se zahraničními partnery, zejména Evropskou environmentální agenturou (EEA) a zástupci zemí V4, kteří výsledky průběžně konzultují a částečně přebírají. Z EEA jsou řešitelům poskytovány informace týkající se převážně technické specifikace výměnných datových formátů pro mezinárodní komunikaci. Všechny vytvořené systémy, standardy a doporučení, které jsou výstupy těchto projektů, reflektují požadavky ve směru informování veřejnosti podle nových evropských předpisů. Zároveň ctí informační pyramidu MDIAR (Monitoring - Data - Informace - Assessment - Reporting) definovanou EEA. Výstupy poskytují informační základnu i pro strategické plánování ve spolupráci s komerční sférou.

LITERATURA

- [1] Barbour, M.T., Yoder, C.O. (2000): The Multimetric Approach to Bioassessment, as Used in the United States of America. In: *Assessing the Biological Quality of Fresh Waters. RIVPACS and Other Techniques*. Freshwater Biological Association, Ambleside, Cumbria.
- [2] Barbour, M.T. (et al.) (1999): *Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates and Fish*, Second Edition. EPA 841-B-99-002. U.S. Environmental Protection Agency, Office of Water, Washington.
- [3] Clarke, R.T. (et al.) (2003): *RIVPACS Models for Predicting the Expected Macroinvertebrate Fauna and Assessing the Ecological Quality of Rivers*. Ecological Modelling.
- [4] Directive 2000/60/EC: *Establishing a Framework for Community Action in the Field of Water Policy*.
- [5] Hughes, R.M. (1994): *Defining Acceptable Biological Status by Comparing with Reference Conditions*. In: Davies, W. S., Simon T. P. (eds.): *Biological Assessment and Criteria. Tools for Water Resource Planning and Decision Making*. Lewis Publishers, Boca Raton.
- [6] Legendre, P., Legendre, L. (1998): *Numerical Ecology*. Elsevier Science BV, Amsterdam.
- [7] Logan, P., Furse, M. (2002): *Preparing for the European Water Framework Directive - Making the Links between Habitat and Aquatic Biota*. *Aquatic Conservation-Marine and Freshwater Ecosystems* 12.
- [8] Rumbaugh, J., Jacobson, I., Booch, G. (2004): *The Unified Modelling Language Reference Manual*, 2nd ed., Addison-Wesley, Boston.
- [9] Washington, H. G. (1984): *Diversity, Biotic and Similarity Indices. A Review with Special Relevance to Aquatic Ecosystems*. *Water Research* 18.
- [10] Wright, J.F. (et al.) (2000): *Assessing the Biological Quality of Freshwaters: RIVPACS and Similar Techniques*. Freshwater Biological Association.

Príspevek vznikl za podpory projektu ARROW (Assessment of Reference Reports of Water Monitoring) Ministerstva životního prostředí ČR a výzkumného záměru INCHEMBIOL (Interakce mezi chemickými látkami, prostředím a biologickými systémy a jejich důsledky na globální, regionální a lokální úrovni, č. MSM0021622412) Ministerstva školství, mládeže a tělovýchovy ČR.